Particles in Water

Properties and Processes

John Gregory

IWA Publishing

CRC Taylor & Francis
Taylor & Francis Group

Particles in Water

Properties and Processes

Particles in Water

Properties and Processes

John Gregory

University College London
England

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.

Co-published by IWA Publishing, Alliance House, 12 Caxton Street, London SW1H 0QS, UK
Telephone: +44 (0)20 7654 5500; Fax: +44 (0)20 7654 5555; Email: publications@iwap.co.uk
Web: http://www.iwapublishing.com/
ISBN: 1-84339-102-3 (Print Edition) (IWA Publishing)

Published in 2006 by CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487–2742

CRC Press is an imprint of Taylor & Francis Group
This edition published in the Taylor & Francis e-Library, 2005.

ISBN 0-203-50845-9 Master e-book ISBN

ISBN 0-203-62148-4 (OEB Format)
International Standard Book Number-10: 1-58716-085-4 (Print Edition) (Hardcover)
International Standard Book Number-13: 978-1-58716-085-1 (Print Edition) (Hardcover)

Library of Congress Card Number 2005041866

Library of Congress Cataloging-in-Publication Data
Gregory, J. (John), 1938– Particles in water: properties and processes/John Gregory. p. cm. Includes bibliographical references and index. ISBN 1-58716-085-4 1. Water chemistry. 2. Particles. I. Title. GB855.G74 2005 541′.34514–dc22 2005041866

Taylor & Francis Group is the Academic Division of T&F Informa plc.
Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com
and the CRC Press Web site at http://WWWcrcpresscom

Dedication

To my wife, Dzintra, for her tolerance and support over the many years of my involvement with the subject of particles in water.

Preface

My professional involvement with particles in water began more than 40 years ago, with studies for my Ph.D. (University of London, 1964) on the role of colloid interactions in deep bed filtration. Since then, my research has been concerned mainly with various aspects of water treatment, especially the removal of particles from water. This book is a distillation of a great deal of experience in the area of aquatic particles, from a mainly fundamental standpoint but with some attention to more practical aspects.

Although I have a degree in chemistry, I had no f ormal training in colloid science and am entirely self-taught in this subject. For this reason, I have tended to acquire knowledge in a rather selective and piecemeal manner, without straying too far from topics directly relevant to my research. Although this strategy has its limitations, the experience has given me a good appreciation of difficulties faced by those just entering the field, and I have tried to present material in an easily understandable manner, without going into a lot of technical detail. In particular, although important mathematical results are presented and their implications are discussed, very few derivations are given.

This approach has some similarity to one advocated by Sir Walter Scott. In the "Introductory Epistle" to his novel *The Abbot*, (1820), Sir Walter writes: "But, after all, it is better that the travellers should have to step over a ditch than to wade through a morass—that the reader should have to suppose what may easily be inferred than be obliged to creep through pages of dull explanation." Although there could be some doubt in several cases about "what may easily be inferred," I hope that my readers will not mind taking convenient shortcuts.

Chapter one is a brief introduction, outlining the origin, nature, and properties of particles in water. This is followed by Chapter two, which deals with particle size, transport processes, and light scattering. There is also a brief section dealing in broad outline with important techniques for particle size determination. Chapter three covers the important topic of surface charge, which plays a major role in colloid stability. Interactions between particles ("colloid interactions") and colloid stability are discussed in Chapter four, with some emphasis on the role of dissolved salts. Chapter five gives an account of particle aggregation kinetics, the form of aggregates, and aggregate strength, all of which are of considerable fundamental and prac-tical importance. Chapter six deals with coagulation and flocculation and the modes of action of some common additives that are used in these processes. The book concludes with Chapter seven, which gives an overview of some important solid-liquid separation processes and the principles on which they are based.

Much of the material is applicable to rather dilute suspensions of particles and is relevant to water treatment processes. However, many of the principles discussed are of more general application and should be of some interest to all who deal with aqueous suspensions. Some knowledge of basic chemistry, physics, and mathematics is assumed.

I am grateful to many colleagues and students around the world, who have provided much needed intellectual stimulation from time to time. Special thanks go to my long-term colleague, friend, and Ph.D. advisor, Professor Ken Ives, who introduced me to the subject of particles in water and gave me the opportunity to pursue research in this area.

John Gregory

About the author

John Gregory is Emeritus Professor of Water Chemistry, Department of Civil and Environmental Engineering, University College London. He has a B.Sc. (chemistry) and a Ph.D. (physical chemistry), both from the University of London.

Although nearly all of his professional life has been spent at University College London, Professor Gregory has had sabbatical periods at universities in Pittsburgh and Delaware (United States), Karlsruhe (Germany), and Perth (Australia).

Professor Gregory has about 40 years of experience in teaching and research in the areas of water pollution and water treatment. His research work has f ocused mainly on physicochemical treatment processes, especially flocculation and filtration, and he has authored more than 100 publications in these and related areas. He is internationally known for his research on polymeric flocculants, colloidal interactions, and monitoring techniques. He introduced the well-known “electrostatic patch” model for flocculation by polyelectrolytes in 1973. During the 1980s, he was involved in the development of a simple monitoring technique for particles and aggregates in flowing suspensions, which has been commercialized and is widely used around the world. He has been invited to lecture on his research at many international conferences.

Professor Gregory has served in various capacities on several professional bodies, including as Advisor to the AWWA Research Committee on Coagulation and Chairman of the Filtration Society (United Kingdom). He has been on the Council of the International Association of Colloid and Interface Scientists (IACIS) and was an Associate Member of the IUPAC Commission on Colloid and Surface Chemistry He is European Editor of *Environmental Engineering Science* and serves on the editorial boards of *Aqua* and *Colloids and Surfaces.*

Contents

chapter one
Introduction

1.1 Particles in the aquatic environment

1.1.1 Origin and nature

Natural waters contain a wide range of impurities, mostly arising from weathering of rocks and soils *(runoff)*. Contributions from human activities, especially domestic and industrial wastewaters, can also be important. Aquatic life is also a significant source of numerous constituents of natural waters.

The most fundamental point concerning impurities in the aquatic environment is the distinction between *dissolved* and *particulate* forms. Although this is a simple concept in principle, distinguishing between these forms is not always straightforward (see later in this chapter). Water is a good solvent for many substances, especially inorganic salts, and the vast majority of dissolved impurities in natural waters are of this type. The *total dissolved solids (TDS)* value for most fresh waters is in the range of 50–1000 mg/L (g/m^3) of which at least 90% would normally be dissolved salts. For sea water, the TDS level is of the order of 35 g/L, most of which is sodium chloride, with other salts making up nearly all of the rest.

Substances that are relatively insoluble in water may exist as small particles, which remain *suspended* for long periods (days or weeks). The total concentration of these impurities is known as the *total suspended solids (TSS)* level. In many natural waters this is very much less than the TDS level, usually of the order of 10–20 mg/L. However, there are important exceptions, especially where heavy seasonal rainf all carries large quantities of particulate material into rivers. For instance, the Yellow River in China can carry up to several g/L of suspended solids. The combined (dissolved+suspended) impurities in water are known as *total solids (TS)*.

The main types of suspended particles found in natural waters are as follows:

- Inorganic
- Organic, including macromolecules
- Living and dead organisms

Inorganic particles result mainly from natural weathering processes and include clays, such as kaolinite and montmorillonite; oxides, including various iron oxides; and silica, calcite, and many other minerals.

Organic material mainly comes from biological degradation of plant and animal remains. Collectively, these substances are known as *natural organic matter (NOM)* and much of this material exists in solution. However, the size of the molecules can be quite large (macromolecules) and they have some characteristics of "particles" (see later in this chapter). Common examples of "true" organic particles are cell debris. NOM in water is usually measured as *total organic carbon* (TOC), and the soluble fraction is measured as *dissolved organic carbon (DOC).* The majority of the DOC falls into the class of material known as *humic substances,* which are responsible for the characteristic peaty-brown color of many natural waters.

Various forms of *aquatic life* could be considered as "particles" in water, but attention is generally restricted to single-celled microorganisms. There is an enormous range of such organisms in natural waters, and, in order of increasing size, they are classified as viruses (although these are not strictly cells), bacteria, algae (including diatoms), and protozoa. Many of these may exist in water either as single cells or as much larger colonies.

Some examples of particles in natural waters are shown as micrographs in Figure 1.1. It is clear from this figure that aquatic particles have sizes over a very wide range. Particle size will be considered in the next section.

In many cases, particles are aggregated to some extent (see Figure 1.1c) so that the apparent size is larger than that of the individual particles. Aggregation of particles will be dealt with in some detail in later chapters. In the marine environment, large aggregates can form, which are known as "marine snow." These aggregates can be several millimeters in diameter and are made up of microorganisms such as bacteria, diatoms, fecal pellets, and many other components, bound together by polysaccharide tendrils. An example is shown in Figure 1.1d. Marine snow settles rapidly in the oceans and is thought to play a major role in transporting carbon and other nutrients to the sea floor.

1.1.2 Particle size ranges

The diagram in Figure 1.2 has a logarithmic scale of particle size over a range of 7 orders of magnitude from 1 mm down to 1 Ångstrom unit (1 Å=10^{-10} m). The sizes range from atomic dimensions to those typical of sand grains. The diagram shows the wavelengths of various forms of electromagnetic radiation, together with typical size ranges of microorganisms and inorganic particles, such as clays. (Other inorganic particles such as ferric oxide cover a similar size range).

Particles larger than about 50 μm are usually visible as discrete objects by the unaided eye. Smaller particles have to be viewed microscopically. Down to about 1 μm an ordinary light microscope can be used, but smaller particles are difficult to resolve because they are comparable in size to the

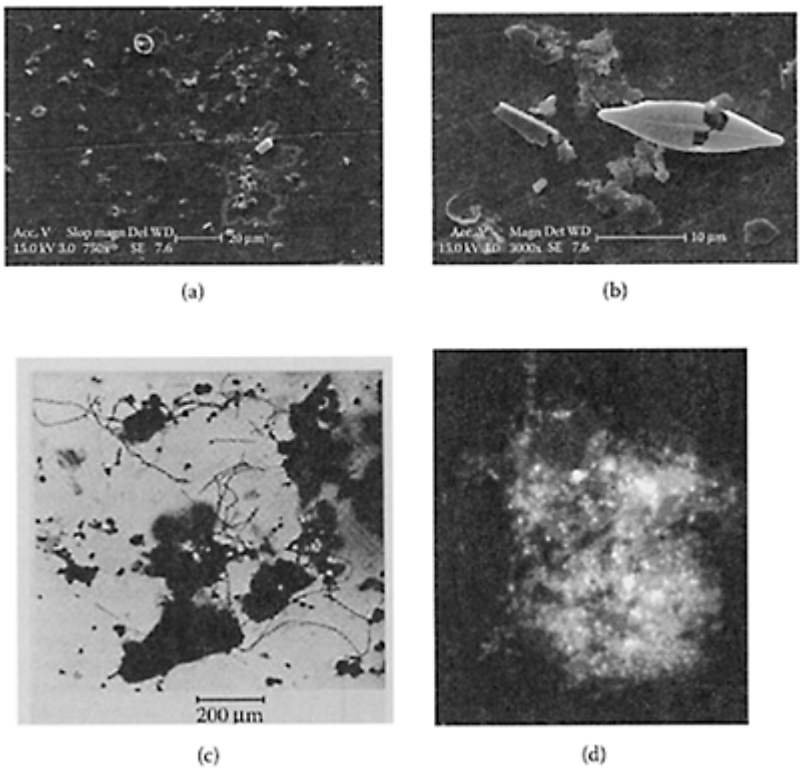

Figure 1.1 Electron micrographs of aquatic particles: (a) Sample from the river Tamar, England. (b) Same source, but higher magnification. (a and b from Doucet, F.J. et al, *J. Environ. Monit.,* 115–121, 2005. With permission). (c) Clay and iron hydroxide particles aggregated by a network of organic filaments. (Note much larger magnification than a and b.) (By permission of Prof. J. Buffle). (d) "Marine snow" from Monterey Bay, CA, mainly colonized by bacteria. The picture shows a 37 µm×27 µm section of a much larger aggregate. (From Azam, F. and Long,

R.A., *Nature,* 414, 495–498, 2001. With permission).

wavelength of visible light. For particles smaller than 1 µm, down to a few nm, the electron microscope provides good resolution because of the much shorter effective wavelength of electrons.

Very small particles (around 20 nm or less) are of similar size to dissolved macromolecules, and in this region the distinction between particles and soluble matter becomes rather vague. In fact, an operational distinction between "dissolved" and "particulate" impurities is often made on the basis of a membrane filtration procedure (*microfiltration;* see Chapter 7). Essentially, the water sample is filtered through a membrane with a defined pore size—often 0.45 µm. Impurities that pass through the filter (in the *permeate*) are conventionally regarded as "dissolved," and those that are retained by the filter (the *retentate*) are defined as "particulate." This is an entirely arbitrary procedure, and it should be clear from Figure 1.2 that this definition

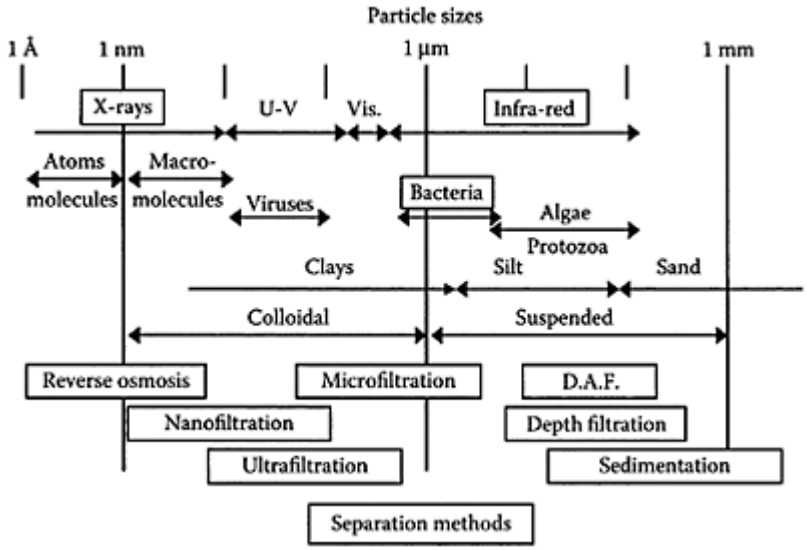

Figure 1.2 Diagram showing range of particle sizes for typical aquatic particles. Also shown are wavelengths of various forms of electromagnetic radiation and appropriate particle separation processes for different size ranges. (D.A.F, dissolved air flotation.)

would put small particles such as viruses and some clays in the "dissolved" category. Nevertheless, a standard procedure has some advantages, and many published listings of dissolved and particulate impurities are based on classification by membrane filtration.

Another very important distinction is between *colloidal* and *suspended* (sometimes called *dispersed*) material. Conventionally the upper limit of the colloidal domain is set at 1 μm, and particles having at least 1 dimension in the size range 1–1000 nm (0.001–1 μm) are known as colloids. There is no sharp change of properties at a particle size of 1 μm, and this is an arbitrary demarcation line. However, there are some important reasons why particles are classified in this way:

- For particles smaller than about 1 μm, diffusion is an important transport mechanism and this tends to prevent particles from settling. Larger particles settle more rapidly and diffusion is less important, so they tend to be removed by sedimentation over time (see Chapter 2).
- At around 1 μm, the *surface area* of particles begins to become significant relative to their volume. For a sphere, the surface/volume ratio is simply $6/d$, where d is the diameter. This means that, for a particle diameter of 1 μm and a density of 2 g/cm^3, the surface area would be 3 m^2/g. For a 10-nm diameter particle the surface area would be 300 m^2/g. A large surface area provides more opportunity for *adsorption* of dissolved impurities (see later in this chapter).
- As particles become smaller, certain types of interaction between particles become more significant relative to "external" forces such as gravity and fluid drag. These *colloid interactions* (see Chapter 4) are important for aggregation and deposition of particles in the colloidal size range.

Although there is no sudden change in properties at a certain particle size, 1 μm represents a convenient boundary, below which these *colloidal phenomena* become more significant. It is also worth noting that, on a logarithmic scale, 1 μm is about halfway between the world of macroscopic objects, such as sand grains, and the molecular domain. A famous book on colloid science by Wolfgang Ostwald, which first appeared in 1914, had the title *Die Welt der vernachlässigten Dimensione (The World of Neglected Dimensions)*. We shall return to the subject of colloids in Section 2 of this chapter.

1.1.3 Effects of particles in water

The presence of particulate impurities in natural waters has a number of important effects, mostly detrimental, on water quality:

- Particles scatter light and give rise to *turbidity* in water (see Chapter 2). Inorganic particles such as clays give a noticeable cloudiness to water at concentrations from just a few mg/L. This is mainly an aesthetic problem, although there may be harmful effects of turbidity on some forms of aquatic life (e.g., predatory fish).
- Because of their relatively high surface area, as mentioned earlier, particles may adsorb certain soluble impurities from water, such as humic substances and trace metals. This can have significant effects on the transport of the adsorbed impurities because they move with the particles. One obvious effect is that dissolved substances, by adsorption on settling particles, can be transported to river, lake, or marine sediments.

• Microorganisms such as viruses and bacteria may be pathogenic (disease-causing) and so the water would be hazardous to human health. Some pathogens may attach to other particles, such as clays, and this could "shield" them from disinfectants used in water treatment processes.

For various reasons, particles are undesirable in drinking water, and the main aim of most common water treatment processes is particle removal (Chapter 7). The boxes in the lower part of Figure 1.2 give the ranges of particle size over which some separation techniques are appropriate.

1.2 Colloidal aspects

Subsequent chapters of this book will go into colloidal phenomena in some detail, but it is worthwhile to make some general points here.

1.2.1 Classification of colloids

The definition of colloids in terms of a size range (1–1000 nm) is convenient, but a further classification is widely used. Since the early days of colloid science, colloids in water have been divided into two distinct groups—*hydrophilic* and *hydrophobic* ("water loving" and "water hating"). For nonaqueous colloids, the corresponding terms are *lyophilic* and *lyophobic* ("liquid-loving" and "liquid-hating"). We are only concerned here with aqueous colloids.

Hydrophilic colloids are essentially water-soluble macromolecules, such as proteins, gums, starch, and many synthetic polymers, generally in the size range 1–10 nm but considerably larger in the case of very-high-molecular-weight polymers. Much of the natural organic matter in water, such as humic substances, can be regarded as hydrophilic colloids. The term "colloid" itself, first coined by Thomas Graham (1805–1869), comes from the Greek word for "glue," so it originally was applied to hydrophilic substances. However, Graham (known as the Father of Colloid Science) also worked with inorganic colloids and the term came to apply to all substances with sufficiently small particles.

A fundamental property of hydrophilic colloids is that they have an affinity for water and are *stable* in a thermodynamic sense. Unless subject to chemical or biological change, they should remain in solution indefinitely. This stability can be reduced by lowering the solubility of colloids in water—for instance, by changing the temperature, as in the "coagulation" of protein in egg white by heating. Changes in chemical conditions, such as pH value, can also cause destabilization (precipitation) of hydrophilic colloids. In many cases, a large increase of salt concentration can cause precipitation *(salting out),* especially of proteins.

Solutions of hydrophilic colloids *(sols),* even at only moderate concentrations, can have markedly different properties from water, mainly because of the large size of the molecules. The most obvious example is the high *viscosity* of many polymer solutions, owing to the extended nature of the polymer chains. The surface tension of macromolecular solutions may also be lower than that of water because of adsorption at the air/water interface.

Although hydrophilic colloids are in true solution, they have some characteristics of "particles," simply because of their large molecular size. For instance, they can scatter light and are unable to pass through dialysis membranes, unlike smaller molecules.

Hydrophobic colloids are substances that are insoluble in water, but they are dispersed as very small particles. Typical examples are inorganic materials, such as clays and oxides, which may be present in water as particles over a wide size range (see Figure 1.2). The term "hydrophobic" is misleading because it normally ref ers to materials that are not easily wetted by water, such as Teflon and talc. Surfaces of these materials show a finite *contact angle* with water, whereas hydrophilic substances are fully wetted, with a contact angle of zero. (This topic is relevant to the removal of particles by flotation; see Chapter 7.) In the colloid context, "hydrophobic" simply means that the material is insoluble in water, irrespective of wetability. Indeed, truly hydrophobic particles are difficult or impossible to disperse in water because they cannot be fully wetted.

The most important difference between hydrophobic and hydrophilic colloids is that the former are *not* thermodynamically stable. The interface between a particle and water has a characteristic *interfacial energy*. The easiest way to envisage this is to consider that work is needed to divide a lump of matter into smaller particles, which is manifested as surface or interfacial energy. The smaller the particles, the larger the total surface area and hence the larger the interfacial energy. This means that small particles have a greater interfacial energy per unit mass than larger particles and would achieve a more stable (lower energy) state by aggregating with other particles to reduce the area in contact with water. The fact that hydrophobic colloids can remain in a finely dispersed state for very long periods is the result of a *kinetic* stability because particle contacts are hindered by repulsive forces. The subject of colloid stability will be discussed briefly in the next section and again, in more detail, in Chapter 4.

The distinction between hydrophilic and hydrophobic colloids in natural waters is often blurred because dissolved organic substances can adsorb on inorganic particles so that the latter may acquire some hydrophilic character. It has been found that particles in the marine environment, although of widely different nature, have similar surface properties, such as zeta potential (see Chapter 3). This is because particles become coated by a layer of organic material so that their surface properties are characteristic of the organic layer, rather than of the underlying particles.

1.2.2 Stability of hydrophobic colloids

Despite their thermodynamic instability, hydrophobic colloids can remain suspended as individual particles for very long periods. Particles in water collide with each other (see Chapter 5) and have ample opportunity to form aggregates. The fact that this does not happen in many cases is the result of forces of repulsion between particles, which keep them from coming into true contact. The most usual reason is that aquatic particles nearly always have a *surface charge,* giving electrical repulsion between particles. Surface charge is the subject of Chapter 3, and its effect on colloid stability will be considered in Chapter 4. Other possible sources of repulsion between particles will also be discussed in Chapter 4.

When electrical repulsion is the origin of colloid stability, then it can be greatly influenced by dissolved salts in water (see Chapter 4). There are many ways of reducing

colloid stability so that aggregation of particles can occur. Particle aggregation may be variously known as *coagulation or flocculation,* and the additives used may be called *coagulants* or *flocculants* (see Chapter 6). These processes are widely used commercially as an essential part of *solid-liquid separation.* Particle separation processes are considered in Chapter 7, but they will be briefly introduced in the next section.

1.2.3 Particle separation processes

Particles in water may be undesirable impurities that have to be removed, as in water and effluent treatment, or they may be valuable materials that need to be recovered, as in mineral processing or biotechnology.

Essentially, particles may be removed from water by the following methods:

- Sedimentation (including centrifugal methods)
- Flotation (including dispersed air and dissolved air methods)
- Filtration (including deep bed and membrane filtration)

All of these processes are greatly dependent on the size of the particles to be removed, and it is often necessary to increase particle size, by coagulation/flocculation processes, to give effective separation. Colloid interactions are vitally important—in particle aggregation and by their effect on the adhesion of particles to other surfaces (such as filter grains and air bubbles).

Many of the topics covered in the following chapters are fundamentally important in particle separation processes. Most of the emphasis will be on fairly dilute suspensions, typical of those encountered in water treatment processes, but the basic principles apply to solid-liquid separation in a wide range of industries, including biotechnology, mineral processing, papermaking, and others.

Further reading

Tadros, Th.F. and Gregory, J., (Eds.), *Colloids in the Aquatic Environment,* Elsevier Applied Science, London. (Special issue of *Colloids and Surfaces A* 73, 1993.)

Wotton, R.S. *The Biology of Particles in Aquatic Systems,* CRC Press, Boca Raton, FL, 1994.

chapter two
Particle size and related properties

2.1 Particle size and shape

"Particles" in water may range in size from a few nanometers (macromolecules) up to millimeter dimensions (sand grains). Natural particles also have various shapes, including rods, plates, and spheres, with many variations in between, which make a treatment of particle size difficult.

The discussion is vastly simplified if the particles are considered to be spherical. In this case, only one size parameter is needed (the diameter) and hydrodynamic properties are much more easily treated. Of course, nonspherical particles are of great importance in natural waters and some way of characterizing them is essential. A common concept is that of the "equivalent sphere," based on a chosen property of the particles.

For instance, an irregular particle has a certain surface area and the equivalent sphere could be chosen as that having the same surface area. The surface area of a sphere, with diameter d, is just πd^2. So, if the surface area of the nonspherical particle is known, the equivalent spherical diameter can easily be calculated. For an object of a given volume, the sphere has the minimum surface area and so the volume (or mass) of a given particle must be equal to or less than that of the equivalent sphere.

Another common definition of equivalent spherical diameter is based on sedimentation velocity. See Section 2.3.3). In this case, from the sedimentation velocity and density of a particle, the diameter of a sphere of the same material that would settle at the same rate can be calculated. This is sometimes called the "Stokes equivalent diameter."

In what follows, we mainly will deal with the properties of spherical particles, which makes the discussion much simpler. Although real particles are usually not spherical, their behavior can often be approximated in terms of equivalent spheres.

2.2 Particle size distributions

2.2.1 General

Only in special cases are particles in a given suspension all of the same size. An example would be monodisperse latex samples, which are often used in fundamental studies and

specialized applications. In the natural aquatic environment and in practical separation processes, we have to deal with suspensions covering a wide range of particle sizes. In such cases, it is convenient to be able to describe the distribution of particle size in a simple mathematical form. There are many distributions in use for different applications, but we shall only consider a few representative examples.

Generally, a particle size distribution gives the f raction of particles within a defined size range in terms of a *probability or frequency function f(x),* where *x* is some measure of the particle size, such as the diameter. This function is defined so that the fraction of particles in the infinitesimal size interval between *x* and *x+dx* is given by *f(x)dx*. The fraction of particles between sizes x_1 and x_2 is then given by the following:

$$\int_{x_1}^{x_2} f(x)dx$$

There are standard relationships giving the *mean* size, *x,* and the *variance,* σ^2 (where σ is the *standard deviation*):

$$\bar{x} = \int_0^\infty xf(x)dx \qquad (2.1)$$

$$\sigma^2 = \int_0^\infty (x-\bar{x})^2 f(x)dx \qquad (2.2)$$

It is often more convenient to think in terms of a *cumulative distribution function,* F(x), which is the fraction of particles with a size less than x. This is given by the following:

$$F(x) = \int_0^x f(x)dx \qquad (2.3)$$

When expressed as a percentage, this is often referred to as “% undersize.” Because there must be some upper limit to the particle size, it follows

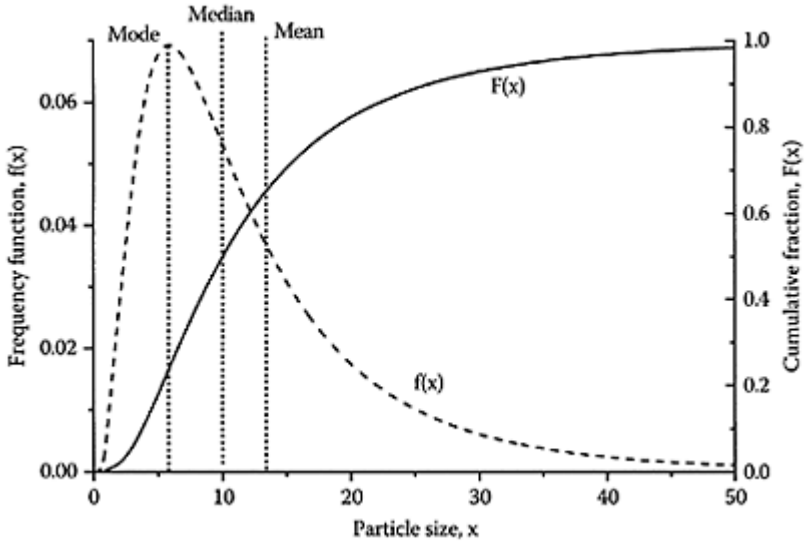

Figure 2.1 Frequency function and cumulative distribution, showing important parameters. The distribution shown is log-normal, Equation (2.11), with median, x_g= 10 and log standard deviation, In σ_g=0.75.

that F(∞)=1. The form of the frequency function f(x) must be such that this condition is satisfied. (The frequency function is then said to be *normalized*.)

Another relationship between the frequency function (or *differential distribution*) and the cumulative distribution is as follows:

$$\frac{dF(x)}{dx} = f(x) \tag{2.4}$$

It follows that the slope of the cumulative distribution *F(x)* at any point is the frequency function *f(x)* at that point. The relation between f(x) and F(x) is shown in Figure 2.1. The maximum in the frequency function (i.e., the most probable size) corresponds to the maximum slope (point of inflection) of the cumulative distribution. For a distribution with just one peak, this is called the *mode* of the distribution, and the distribution is said to be *monomodal*. The *median* size is that corresponding to 50% on a cumulative distribution—that is, half of the particles have sizes smaller (or larger) than the median. The mean size has already been defined by Equation (2.1). For a symmetric distribution, the mean, median, and mode sizes are all the same, but these may differ considerably for an asymmetric distribution (as in Figure 2.1).

So far, our discussion has been in terms of the fractional *number* of particles within a given size range, but there are other ways of presenting particle size distributions. The

most common alternative is the *mass* (or *volume*) distribution, by which the fraction of particle mass or volume within certain size limits can be expressed. For particles of the same material, mass and volume distributions are effectively the same because mass and volume are directly linked through the density of the material. For a mixture of particles of different types, there is no simple relation between mass and number distributions. For simplicity, we shall only consider particles of the same material.

For a sphere, the mass is proportional to the cube of the diameter and this makes a huge difference to the shape of the size distribution. Expressed in terms of particle mass, the differential distribution is as follows:

$$f_m(x)=Bx^3f(x) \tag{2.5}$$

where *B* is a constant that normalizes the function, so that the integral over all particle sizes has a value of unity:

$$\int_0^\infty f_m(x)dx = B\int_0^\infty x^3 f(x)dx = 1 \tag{2.6}$$

This simply states that all particles must have a mass between zero and infinity. (The number frequency function *f(x)* is defined so that $\int_0^\infty f(x)dx = 1$). Size distributions for the same suspension, based on number and mass, are shown in Figure 2.2. The mass distribution is much broader and has a peak (mode) at a considerably larger size. The mean size on a mass basis, $\overline{x_m}$, (often called the "weight average") is given by the following:

$$\overline{x_m}^3 = \int_0^\infty x^3 f(x)dx \tag{2.7}$$

Only for a truly monodisperse suspension would the number and weight averages coincide. The ratio of these values is one measure of the breadth of a distribution.

All of our discussion so far has been in terms of *continuous* distributions −particle size is treated as a parameter that can take any value and *f(x)* is a continuous function of *x*. There are cases where it is more convenient to think in terms of *discrete* distributions. For instance, a suspension may contain known concentrations of particles in discrete size ranges and the distribution can be plotted in the form of a *histogram*, as in Figure 2.3. In fact, experimental methods of determining particle size usually give results in this form. It should be clear from Figure 2.3 that, as the width of the chosen

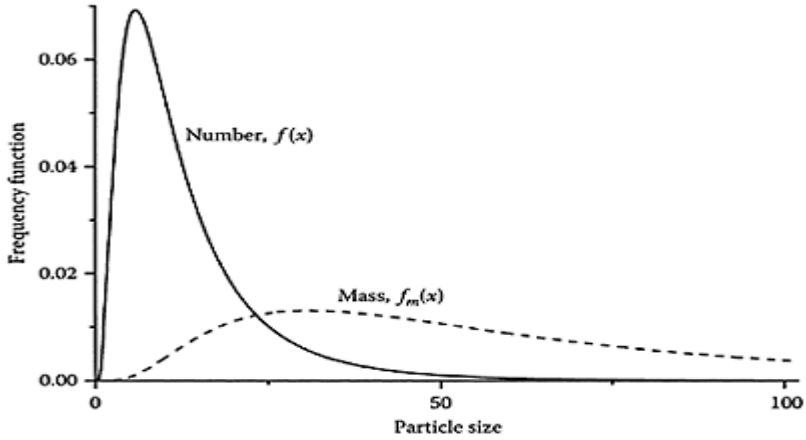

Figure 2.2 Showing number and mass frequency functions for the same distribution as in Figure 2.1.

size interval decreases, the histogram would approach the shape of a continuous distribution. If a mean size is assigned to each interval, then we can say that there are N_1 particles of size x_1, N_2 of size x_2, and so on, with the total number N_T being given by:

$$N_T = \sum_i N_i \ ,$$

where N_i is the number of particles of size x_i and the sum is taken over all the measured sizes.

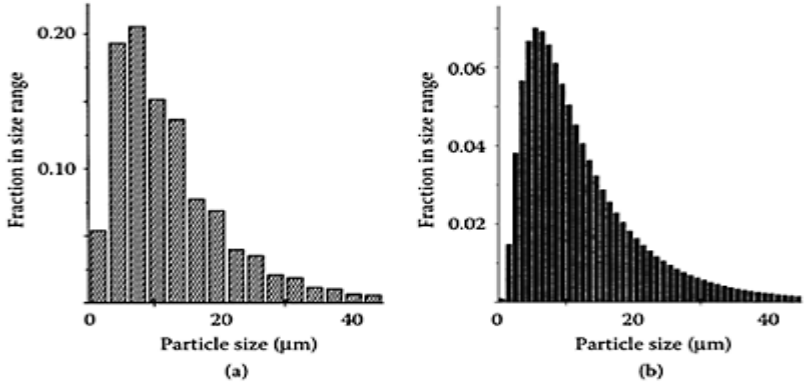

Figure 2.3 Discrete particle size distributions plotted as histograms. Size intervals: (a) 3 μm; (b) 1 μm. Data

derived from the log-normal distribution in Figure 2.1.

The mean size and variance for a discrete distribution are given by the following:

$$\bar{x} = \frac{\sum x_i N_i}{N_T} \tag{2.8}$$

$$\sigma^2 = \frac{\sum N_i (x_i - \bar{x})^2}{N_T} \tag{2.9}$$

These equations are the analogs of Equations (2.1) and (2.2) for continuous distributions.

It is convenient when a particle size distribution, *f(x),* can be represented by a simple mathematical form, since data presentation is then much easier. For instance, the whole distribution could be characterized by only a few parameters, rather than having to report numbers in many different size intervals. In some cases only two parameters are needed, typically a mean size and the standard deviation. Two common forms of particle size distribution will be discussed in the next sections.

2.2.2 The log-normal distribution

Many natural phenomena are known to follow the *normal* or *gaussian* distribution, which, for a variable *x,* can be written as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left[\frac{x - \bar{x}}{\sqrt{2}\sigma} \right]^2 \right) \tag{2.10}$$

where $\bar{x}$ and σ are the mean and standard deviation.

This gives the well-known bell-shaped curve, which is symmetric about the mean. About 68% of the values lie within one standard deviation of the mean and about 95% lie within two standard deviations of the mean $(x \pm 2\sigma)$. Although the normal distribution does not give a good representation of real particle size distributions, a simple modification gives a useful result.

By considering the natural log of the size, *In x,* rather than *x,* as the variable, we arrive at the *log-normal distribution:*

$$f(x) = \frac{1}{\sqrt{(2\pi)}x \ln \sigma_g} \exp\left\{ -\left[\frac{\ln x - \ln \bar{x}_g}{\sqrt{2} \ln \sigma_g} \right]^2 \right\} \tag{2.11}$$

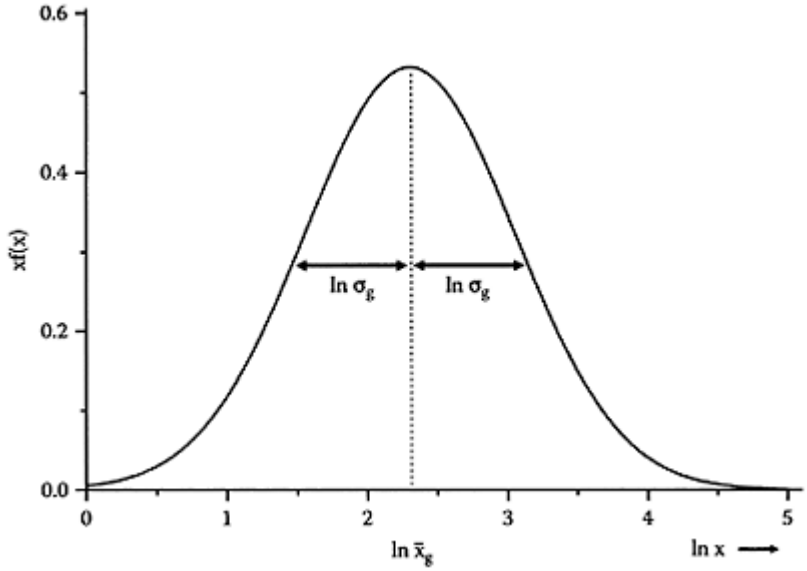

Figure 2.4 *Log*-normal distribution plotted as xf(x) versus In *x*.

In this expression, $\overline{x}_g$ is the geometric mean size (i.e., ln $\overline{x}_g$ is the mean value of ln *x*) and ln σ_g is the standard deviation of In *x*, as follows:

$$\left(\ln \sigma_g\right)^2 = \int_0^\infty \left(\ln x - \ln \overline{x}_g\right)^2 f(x)dx \qquad (2.12)$$

The nature of the log-normal distribution is such that *xf(x)*d(ln*x*) is the fraction of particles with In (size) in the range In *x*−ln *x*+d(ln *x*). It follows that when xf(x) is plotted against In x, the familiar bell-shaped gaussian curve is obtained, as in Figure 2.4. However, when f(x) is plotted against x, a distinct positive skew is apparent, especially for fairly high values of the logarithmic standard deviation. (In fact, the distribution in Figure 2.1 is log-normal).

The log-normal form appears to fit some actual particle size distributions quite well. When plotted on log-probability paper, the cumulative log-normal distribution gives a straight line (Figure 2.5), from which useful information may be derived. For instance, the median size (equivalent to the geometric mean size) can be read immediately as the 50% value. The standard deviation can be obtained by reading the values corresponding to 84% and 16% undersize, which represent sizes one logarithmic standard deviation above and below the median, so that

$$2 \ln \sigma_g = \ln x_{84} - \ln x_{16} \qquad (2.13)$$

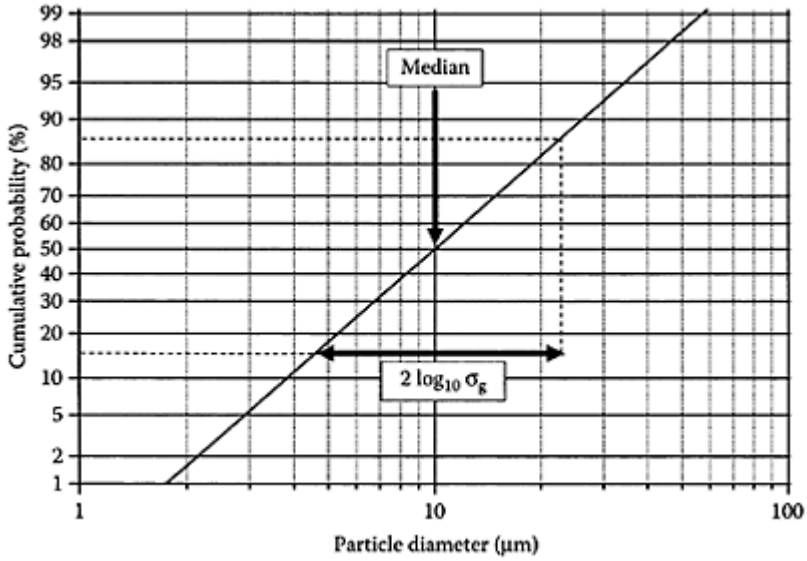

Figure 2.5 Cumulative log-normal distribution plotted as log probability graph (see text).

(Note: when the log-probability paper has a $\log_{10}$ axis, as shown, the value derived by this procedure is $2 \log_{10} \sigma_g$).

One useful feature of the log-normal distribution is that if a sample follows this form on the basis of, say, number concentration, then the distributions on any other basis (e.g., mass or volume) will all have the same form. On a log-probability plot they will give parallel straight lines.

There are some simple relationships for the log-normal distribution that can be useful. For instance, the mean x and mode x_{max} sizes are related to the geometric mean size $\overline{x_g}$ and $\ln \sigma_g$ by the following expressions:

$$\ln \bar{x} = \ln \overline{x_g} + \frac{(\ln \sigma_g)^2}{2} \qquad (2.14)$$

$$\ln x_{max} = \ln \overline{x_g} - (\ln \sigma_g)^2 \qquad (2.15)$$

(The median size is, by definition, equal to $\overline{x_g}$)

Also, if the relative standard deviation of the distribution is defined as $\sigma_r = \sigma / \bar{x}$, then the logarithmic standard deviation is given by the following:

$(\ln \sigma_g)^2 = \ln(1+\sigma_r^2)$ (2.16)

From these expressions, the following conclusions may be drawn for a log-normal distribution.

1. The mean size is always greater than the median size, by an amount that depends greatly on the logarithmic standard deviation.
2. The most probable size (the mode) is always less than the median and mean sizes.
3. For small values of σ_r, it follows from Equation (2.16) that $\ln \sigma_g \approx \sigma_r$. Thus, for a logarithmic standard deviation of 0.1, the standard deviation would be about 10% of the mean size. In fact, for very narrow distributions, there is little difference between the normal and log-normal forms. However, for larger values of $\ln \sigma_g$, distributions can become broad and highly skewed.

Another useful result for a log-normal distribution is the relation between the total number concentration of particles N_T and the volume fraction, ϕ (the volume of particles per unit volume of suspension):

$$\phi = \frac{4}{3} N_T \pi \exp\left(3 \ln \overline{x_g} + \frac{9(\ln \sigma_g)^2}{2} \right) \qquad (2.17)$$

Although the log-normal form fits many actual particle size distributions reasonably well, it is of limited use in describing particles in natural waters. In this case an alternative distribution is often applicable.

2.2.3 The power law distribution

For natural particles in oceans and fresh waters a very simple power law distribution has sometimes been f ound to be appropriate, at least over certain size ranges. This can be written in differential form as follows:

$$\frac{dN}{dx} = n(x) = Zx^{-\beta} \qquad (2.18)$$

Here, N is the number of particles with size less than x and Z and β are empirical constants. The differential function $n(x)$ is related to the frequency function f(x) (see Equation 2.4), but the latter cannot be used in the case of the power law distribution. If Equation (2.18) is integrated over the entire range of x values (zero to infinity), it predicts an infinite amount of particulate material, so that the concept of a fraction of particles within a certain size range is not applicable. The power law distribution can only be used over a finite particle size range.

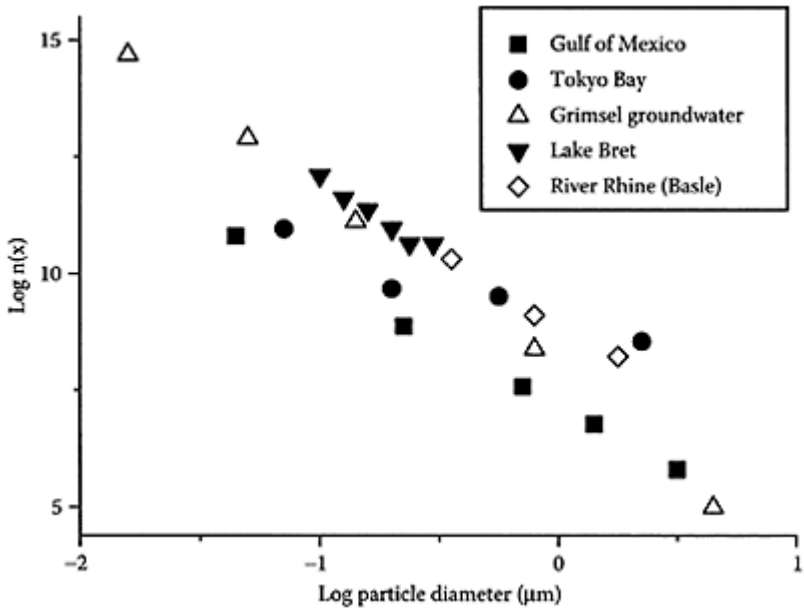

Figure 2.6 Particle size distributions from several natural waters, plotted according to the power law form, Equation (2.18). (Replotted from data in Filella, M. and Buffle, J., *Colloids and Surfaces*, A 73:255–273, 1993.)

The value of the constant Z in Equation (2.18) is related to the total amount of material, and β indicates the breadth of the distribution. Typical values of β for natural waters are in the range 3–5, and mostly around 4. Some examples of size distributions for particles in natural waters are shown in Figure 2.6. They are presented in the form of log-log plots, so that the slope gives the value of β directly. These plots show a continuous increase of *n(x)* as the size decreases, with no peak. (This is why the total particle number approaches infinity as the size goes to zero.) In reality, all reported size distributions are based on particle size measurement techniques (see Section 5 in this chapter), which are limited to certain ranges of size. There is always a lower size limit, beyond which particles cannot be detected, and it is possible that a peak in *n(x)* lies in this inaccessible region. However, it should be remembered (see Figure 1.2) that, in the size range of a few nm and smaller, we are in the realm of dissolved macromolecules as well as smaller molecules and ions, so that a peak in *n(x)* need not exist, if we regard all components, both dissolved and particulate, as "particles."

If the power law distribution is written in terms of the mass (or volume) of particles, then the distribution takes a different form, depending on the value of the exponent β. It is more convenient to write the power law distribution in terms of log (size), thus:

$$\frac{dN}{d(\log x)} = Zxn(x) = Zx^{(1-\beta)} \tag{2.19}$$

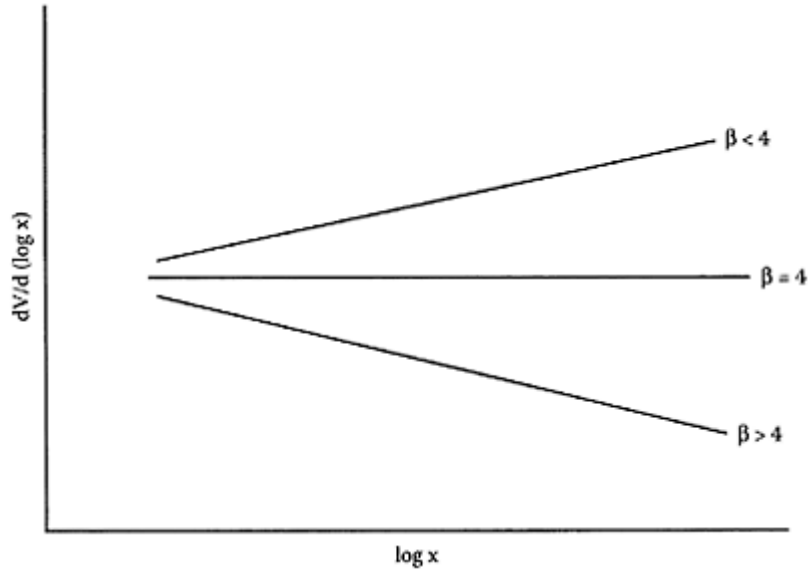

Figure 2.7 Forms of power law distributions on a volume basis, Equation (2.20), for different values of the parameter β.

If we express particle concentration in terms of volume, *V*, the corresponding expression is as follows:

$$\frac{dV}{d(\log x)} = Yx^{(4-\beta)} \tag{2.20}$$

where *Y* is a constant.

For β=4 (a typical value for natural waters), Equation (2.20) implies that the volume of particulate material in a given logarithmic size interval does not depend on the size. Thus, the volume of "particles" in the size range of 1–10 μm (bacteria, algae, etc.) would be the same as that in the range of 1–10 m (sharks, dolphins, whales, etc.). Although it would be difficult to confirm this prediction quantitatively, it illustrates the great difference between number and volume distributions. For other values of the exponent β, different volume distributions emerge, as shown in Figure 2.7. For values of β less than 4 the volume of particles in a given size range increases with size and vice versa.

The simplicity of the power law distribution makes it convenient for describing the fate of particles in the aquatic environment and in water and wastewater treatment plants.

2.3 Particle transport

Particles in water may be transported in various ways, the most significant of which are as follows:

- Convection
- Diffusion
- Sedimentation

Convection is simply the movement of the particles as a result of flow, whereas diffusion and sedimentation cause particles to depart from flow lines. These types of motion are hindered *by fluid drag*.

2.3.1 Fluid drag

Whenever a particle moves in a fluid, a drag force, F_D, is experienced. This is most conveniently presented as a product of two terms: a force term and a dimensionless *drag coefficient*, C_D. The force term is just the dynamic pressure ($=\rho_L U^2/2$) multiplied by an effective area of the particle, *S*. Thus:

$$F_D = \frac{1}{2}\rho_L S U^2 C_D \tag{2.21}$$

The drag coefficient is a function of the Reynolds number, *Re*, given by the following:

$$Re = \frac{\rho_L l U}{\mu} \tag{2.22}$$

ρ_L is the density *U* is the velocity and μ is the viscosity of the fluid and *l* is a characteristic length (for instance, the diameter of a spherical particle). The drag force can then be written as follows:

$$F_D = \frac{1}{2}\rho_L S U^2 f(Re) \tag{2.23}$$

Care has to be taken in choosing an appropriate area, S, because the drag coefficient depends greatly on how S is defined, especially for highly asymmetric shapes, such as fibers. The simplest procedure, and the one adopted here, is to define S as the projected area of the particle on a plane normal to the direction of flow. So, for a sphere, with diameter *d*, $S=\pi d^2/4$.

The dependence of drag coefficient on the Reynolds number is of paramount importance. At low values of *Re*, viscous effects dominate over inertial effects and flows are orderly and *laminar*. As *Re* increases (for instance, either by increasing fluid velocity or the size of the particle) inertial forces become more important, leading to vortices and, eventually, to *turbulence*.

For very low Reynolds numbers (Re<0.1), the so-called “creeping flow” conditions apply and the drag coefficient for a sphere takes a simple form, first derived by Stokes:

$$C_D = \frac{24}{Re} \tag{2.24}$$

This expression is good enough for most particles in natural waters, but in those cases where the Reynolds number is too large for the Stokes expression to apply, the drag coefficient has to be obtained in a different manner. There is no fundamental theory that gives the drag coefficient under non-creeping flow conditions. Instead, we have to rely on experimentally determined values and empirical equations. For spheres, and *Re* values in the range of 1–100, the following expression gives a good approximation to the drag coefficient:

$$C_D = \frac{24}{\mathrm{Re}} + \frac{6}{1 + \mathrm{Re}^{1/2}} + 0.4 \tag{2.25}$$

Drag coefficients calculated from Equation (2.25) are plotted in Figure 2.8, together with the results for "creeping flow" conditions from Equation (2.24). This shows that the departure from the simple form becomes significant at higher Reynolds numbers and that the drag coefficient is higher than that given by assuming very slow flow. The form of the drag coefficient becomes more complicated for much higher *Re* values, but these are not relevant to aquatic particles.

Drag on particles in water has a major influence on the processes of diffusion and sedimentation.

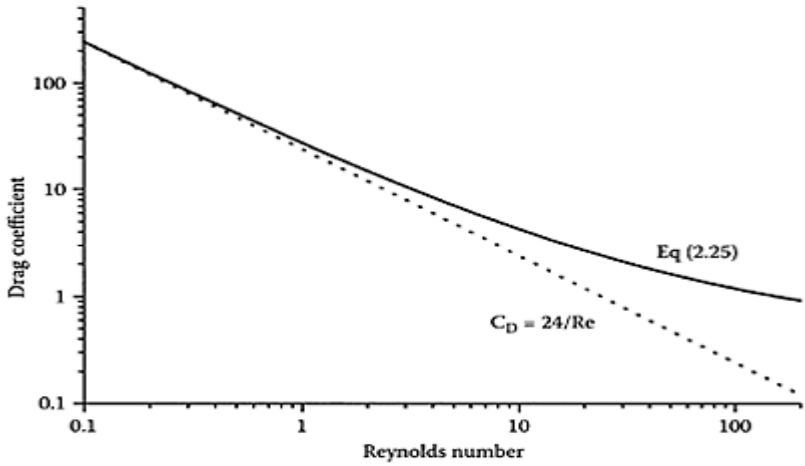

Figure 2.8 Drag coefficient for spheres as a function of the Reynolds number. The full line shows values from Equation (2.25), which are close to actual data. The dashed line shows the

Stokes result for “creeping flow” conditions.

2.3.2 Diffusion

When particles of a few microns in size or less are observed by microscope, it is immediately apparent that they are in constant random motion. There is an endless “jiggling” of particles, which is known as *brownian motion,* after Robert Brown (1773–1858), a British botanist who first reported the effect in 1827. It is a common misconception that Brown observed the motion of pollen grains; these are too large (typically around 20 μm) to show significant brownian motion. This has cast doubt on whether Brown actually observed genuine brownian motion. However, it is clear from Brown’s account that what he actually observed were tiny particles *within* the pollen grains, which were small enough to show the effect.

The origin of brownian motion remained a mystery f or many years until, toward the end of the 19th century, it was realized that the random movement of particles resulted directly from the thermal motion of molecules of the liquid in which they were suspended. In fact, the theoretical treatment of this problem by Einstein and Smoluchowski and the agreement with experimental observations provided the first definitive proof of the long-conjectured existence of atoms and molecules.

The kinetic energy of water molecules means that they are in continuous chaotic motion, and particles suspended in water are constantly being bombarded by these molecules. This imparts kinetic energy to the particles and results in the phenomenon of brownian motion. Although, on average, a particle is just as likely to be struck by water molecules from any direction, it is inevitable that, in a very small time interval, more molecules will strike from one side, giving the particle a small “kick” in one direction. In the next instant the particle may jump in another direction, and so on. The combined effect of these collisions is to make the particle move in a series of steps in random directions, known as a *random walk* or sometimes a “drunkard’s walk.” A two-dimensional version of a random walk is shown in Figure 2.9.

The nature of brownian motion is such that a particle will, on average, move a certain distance from its starting point in a given time. In a time *t,* the mean displacement is given by the following:

$$\bar{x} = \sqrt{2Dt} \tag{2.26}$$

where *D* is the *diffusion coefficient* of the particle.

The diffusion coefficient of a spherical particle, diameter *d,* is given by the *Stokes-Einstein equation:*

$$D = \frac{k_B T}{3\pi d\mu} \tag{2.27}$$

where k_B is the Boltzmann constant and *T* is the absolute temperature.

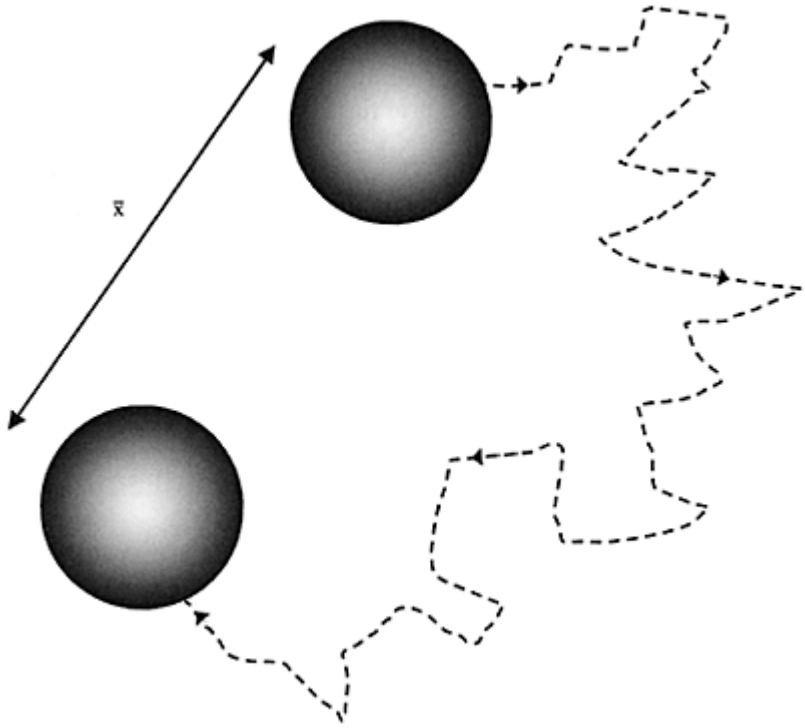

Figure 2.9 Showing a typical "random walk" in two dimensions. In a given time, the particle moves, *on average*, a distance $\bar{x}$ from its starting position.

The numerator in this expression is related to the thermal energy of the particle, which promotes diffusion, and the denominator is a measure of the drag experienced by the particle, which tends to hinder diffusion. It is clear that for larger particles and in more viscous fluids, the diffusion coefficient will be smaller.

As an example, a particle of 1-μm diameter in water at 20°C has a diffusion coefficient of 4.25×10^{-13} m^2/s. Using Equation (2.26), the mean displacement of this particle would be only about 7 μm in 1 minute. The nature of diffusion is such that the distance moved by a particle is proportional to the *square root* of time, so that even after 1 hour a 1-μm particle would only have moved, on average, about 55 μm from its starting position. Nevertheless, diffusion is a significant effect f or very small, colloidal particles and plays an important role in particle aggregation (see Chapter 5).

2.3.3 Sedimentation

Particles in water move in response to gravitational attraction. Again, this motion is retarded by fluid drag, which depends on the particle velocity, according to Equation

(2.23). A particle initially at rest will accelerate until the drag force is exactly balanced by the gravitational force. The particle then moves at a constant speed, known as the *terminal velocity*. Under most conditions the terminal velocity is attained rapidly and we do not need to concern ourselves with the transition period.

The gravitational force depends on the particle volume and the difference in density between the particle and fluid. For a sphere, the force is given by the following:

$$F_g = \frac{\pi d^3}{6}\left(\rho_S - \rho_L\right)g \qquad (2.28)$$

where ρ_s is the density of the particle and ρ_L is the density of the fluid. (If the particle is less dense than the fluid, this equation gives a negative force, which simply means that the particle will move upward or float.)

For creeping flow conditions, where the simple form f or drag coefficient, Equation (2.24), applies, then equating the gravitational and drag forces for a sphere gives the terminal sedimentation velocity:

$$U = \frac{gd^2}{18\mu}\left(\rho_S - \rho_L\right) \qquad (2.29)$$

This is the well-known *Stokes law*, which is widely used in practice. However, it is based on the assumption of very low Reynolds number and hence only applies to low settling rates (i.e., fairly small particles). For larger values of *Re*, Stokes law becomes inaccurate and it is important to establish limits of applicability.

Using the empirical form for C_D, Equation (2.25), and assuming a certain density for particles, it is possible to plot the settling velocity in water against particle diameter, as in Figure 2.10. The particle densities chosen are 2.5 and 1.1 g/cm^3 and corresponding results are shown from Stokes law, Equation (2.29). The assumed viscosity is that for water at 20°C. Because the results are shown as log-log plots, the Stokes result gives straight lines, with slope 2 (because of the dependence of settling rate on the square of particle size). It is clear that significant departures from Stokes law occur for particles above a certain size (around 100 μm and 200 μm for the high- and low-density particles, respectively). The corresponding settling velocities are of the order of 7 and 2 mm/sec. Thus, it should be assumed that if particles are settling at more than a few mm/sec in water, then Stokes law does not apply. In fact, the Stokes result greatly overestimates the actual rate for larger particles—by a factor that can be more than 10 for particles larger than 1000 μm (1 mm). However, for particles in water smaller than about 50 μm, Stokes law will nearly always be adequate.

All of the previous discussion in this chapter has been for *isolated* particles and applies only to rather dilute dispersions. At higher concentrations the motion of a particle is affected by hydrodynamic interaction with nearby particles. This effect leads to the phenomenon of *hindered settling*, which is

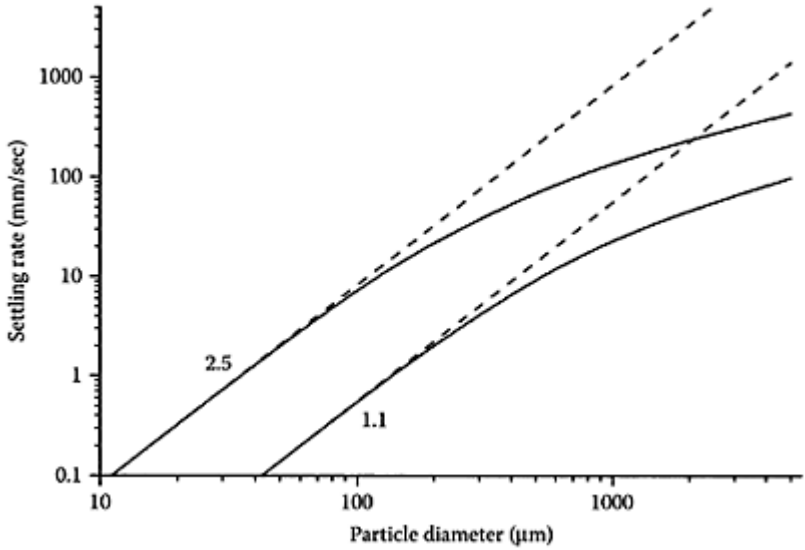

Figure 2.10 Settling rate versus particle diameters for two different particle densities, 1.1 and 2.5 g/cm^3. Full lines: Calculated using drag coefficient from Equation (2.25). Dashed lines: From Stokes law, Equation (2.29).

slower than sedimentation of the isolated particles. Furthermore, hindered settling leads to a condition where all particles in a concentrated suspension settle at the same rate, irrespective of particle size. This is known as *zone settling* and is made visually apparent by a sharp boundary between the sedimenting particles and the clear supernatant liquid. Hindered settling can be important in practice but will not be considered further here.

2.3.4 Effect of particle size

It is useful to compare the relative importance of particle transport by diffusion and sedimentation. These mechanisms differ greatly in their dependence on particle size. Using Equation (2.26) for diffusion together with the Stokes-Einstein diffusion coefficient and Stokes law for sedimentation, we could calculate the time needed for a particle to move through a certain distance (e.g., 1 mm). However, the choice of distance is arbitrary and can give misleading results (because the diffusion distance varies as the square root of time). Another way is to calculate the time during which a particle would move through a characteristic distance, corresponding to its own diameter (e.g., 10 μm for a 10-μm particle). This is the basis of the results in Figure 2.11, where, as in Figure

2.10, the particles are assumed to be spherical with densities of 2.5 and 1.1 g/cm^3 in water at 20°C.

Figure 2.11 shows that the characteristic time for diffusion is much less than that for sedimentation when particles are small. For 0.1-μm particles, the time to diffuse through 0.1 μm is around 10,000 times shorter than the

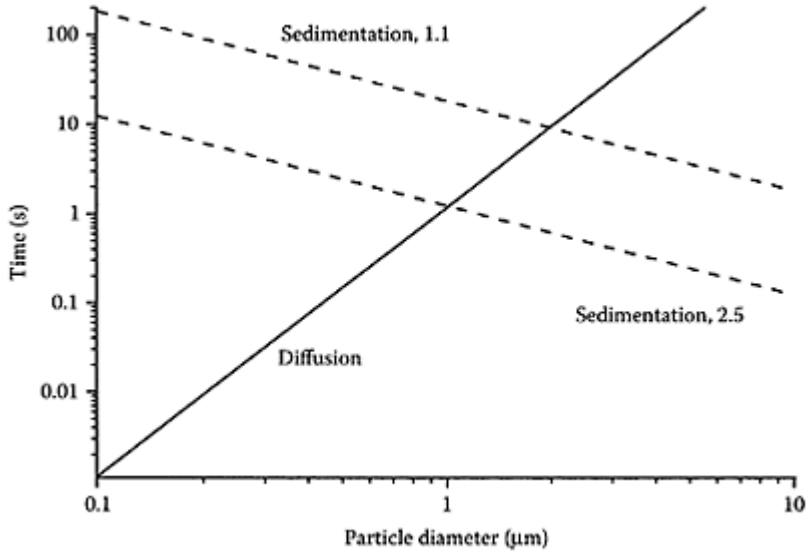

Figure 2.11 Characteristic times for a particle to move through a distance equal to its own diameter, by diffusion and sedimentation. In the latter case, the densities assumed are the same as those for Figure 2.10.

corresponding sedimentation time. However, for 10-μm particles the diffusion time can be several hundred times longer than for sedimentation. There is a crossover point of the lines in Figure 2.11, which depends on the particle density but is generally of the order of 1 μm. This is one reason why 1 μm is conventionally taken as the upper limit of the colloidal size range. Smaller particles tend to remain in suspension because diffusion is more significant than sedimentation. Larger particles tend to settle out over time.

2.4 Light scattering and turbidity

2.4.1 General

When a light beam illuminates a suspension of particles, the intensity of the transmitted beam is reduced as a result of *scattering* and *absorption* of light.

Absorption is a result of radiation of a specific wavelength interacting with atoms and molecules to raise their energy levels. This involves the loss of energy from the beam, which is ultimately dissipated as heat. When visible light is absorbed, the effect is made apparent as a characteristic color. Although absorption by particles in water may be significant in some cases, we shall not consider this effect further. Light scattering is usually a more important effect, and discussion of scattering is made much more complicated if absorption is included.

Light scattering occurs with all particles in water and involves no net loss of energy from the beam. Electromagnetic radiation induces displace-

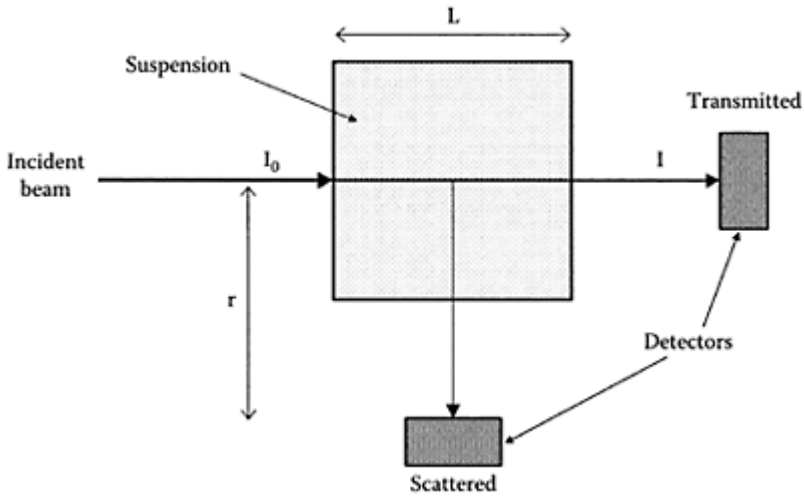

Figure 2.12 Schematic illustration of a setup for measuring transmitted and scattered (90 degrees) light from a suspension of particles.

ments of electrons and hence fluctuating dipoles within particles, which radiate energy in all directions at the same frequency as the incident radiation. The observed scattering behavior is a result of the interference of the radiated and incident light. The intensity of the induced dipoles depends on the *polarizability* of the material and hence on the *refractive index*. The only requirement for light scattering is that the particle has a

refractive index different from that of water. Particles that are completely transparent to the incident radiation (i.e., with no absorption) are effective scatterers, provided that there is a refractive index difference.

Because there is no net loss of energy from the incident beam, the process is often known as *elastic* scattering. However, the transmitted beam shows some loss of intensity because radiation is scattered in all directions and so less would reach a detector placed opposite to the light source. For the same reason, a detector placed at some angle to the light beam would receive more radiation, as a result of light scattered in that direction. These concepts are illustrated in Figure 2.12.

Figure 2.12 shows the two practical methods of measuring *turbidity* of a suspension. The turbidity is a direct consequence of light scattering and can be measured either as a reduction in intensity of transmitted light or an increase in scattered light intensity at a chosen angle (often 90 degrees) to the beam. Turbidity is the most common visible evidence of particles in water. Even at low concentrations, particles can impart a noticeable cloudiness (turbidity) to water.

Light scattering depends on the following particle properties:

- The size of the particles relative to the light wavelength
- The shape of the particles
- The refractive index of the particles relative to that of the suspension medium.

A complete mathematical treatment of light scattering by particles is complex and will not be given here. We shall restrict attention to some simple cases and to spherical particles (so that shape effects are not considered). Initially, the concepts of scattering cross-section and specific turbidity will be introduced.

2.4.2 Turbidity and light transmission

In the arrangement shown in Figure 2.12 the incident light intensity is I_0 and the intensity of light transmitted through the suspension is I. According to the *Beer-Lambert law*, the intensity of transmitted light depends exponentially on the particle concentration and the optical path length, L. It is convenient to assume that all the particles have the same size, although the Beer-Lambert law does not depend on this assumption. If the particle number concentration is N, this law can then be written as follows:

$$I=I_0 \exp(-NCL) \tag{2.30}$$

where C is the *light scattering cross-section* of a particle (see below).

It should be clear from Equation (2.30) that the incident intensity, I_0, is just the intensity of light transmitted through a clean, particle-free fluid (N =0). Thus any light scattering or absorption by the fluid or the optical cell is taken care of.

The scattering cross-section C has units of area and is an important quantity. Essentially, each particle scatters an amount of light energy corresponding to that in an area C of the incident beam. Let us consider a thin layer or “slice” of suspension, with thickness δL and illuminated by a light beam of intensity I (see Figure 2.13). Because the total beam area is A, the volume of the layer illuminated by the beam is just $A\delta L$ and the number of particles in this volume is $NA\delta L$. If each particle effectively “blocks” an area

C, then the fractional area of the beam from which light is removed by scattering is $NCA\delta L/A=NC\delta L$. This is effectively the fraction of light removed from the beam by scattering. Thus the change of intensity as the light passes through the thin layer is given by: $\delta I=-INC\delta L$. In differential form:

$$\frac{dI}{I}=-NCdL \tag{2.31}$$

When this is integrated with $I=I_0$ when $L=0$, we get the following:

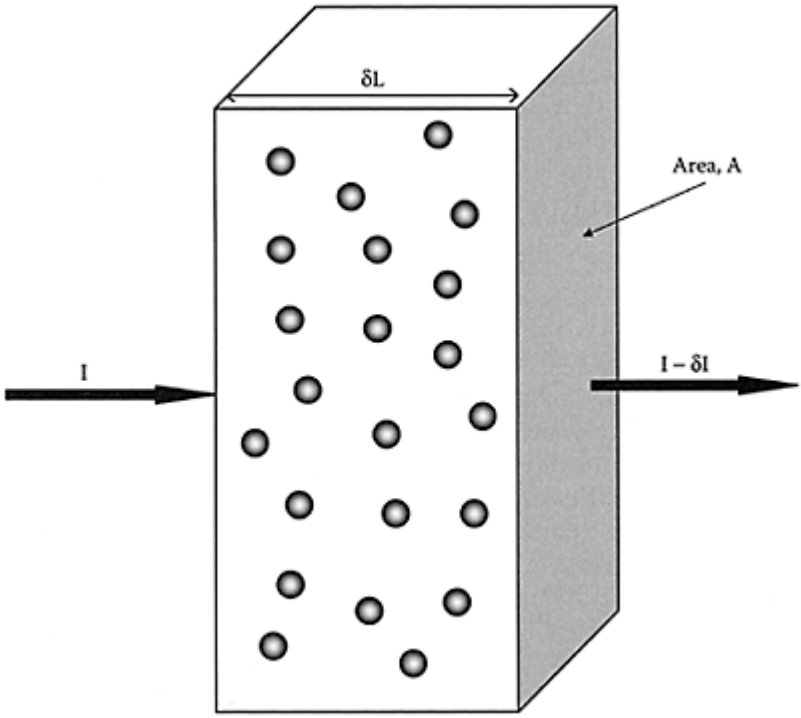

Figure 2.13 Light transmission through a thin layer of suspension (see text).

$$\ln\left(\frac{I}{I_0}\right)=-NCL \tag{2.32}$$

This is just another form of the Beer-Lambert law, Equation (2.30). The basic assumption is that each layer removes the same fraction of light incident on it as all other layers.

The Beer-Lambert law can also be written in terms of the *turbidity*, τ, which is defined as follows:

$$\tau = NC \tag{2.33}$$

This only applies if all particles have the same scattering cross-section. For suspensions with different particles, the corresponding expression is as follows:

$$\tau = \sum N_i C_i \tag{2.34}$$

where N_i is the number concentration of particles with scattering cross-section C_i, etc., and the sum is taken over all types of particles.

(These expressions only apply to turbidity measured by transmission. There are no equivalent definitions of turbidity from measurements of scattered light, and calibration standards have to be used.)

It is reasonable to suppose that the scattering cross-section of a particle depends on its geometric cross-sectional area. The *scattering coefficient Q* is simply the ratio of these areas. Thus, for a sphere, diameter *d*, scattering cross section is given by the following:

$$C = \frac{Q\pi d^2}{4} \tag{2.35}$$

We shall see later that *Q* can vary from close to zero to values much greater than 1, so that a particle may scatter very much less or considerably more light than that incident on it.

For spherical particles of diameter *d*, the volume of each particle is $\pi d^3/6$, so the total volume of particles per unit volume of suspension (the *volume fraction*, ϕ) is as follows:

$$\phi = \frac{N\pi d^3}{6} \tag{2.36}$$

Combining this with Equations (2.33) and (2.35) gives the following:

$$\frac{\tau}{\phi} = \frac{3Q}{2d} \tag{2.37}$$

The term τ/ϕ, the turbidity normalized by the particle volume concentration, is called the *specific turbidity* and it is clear from Equation (2.37) that this varies inversely with particle size. However, it is also proportional to the scattering coefficient Q and we must now consider how this quantity varies with particle size.

Light scattering theory is complex and we cannot go into details here. The full theory, usually attributed to Gustav Mie, gives information on the angular distribution of scattered light from a spherical particle, as well as on the total scattering (and hence the scattering coefficient). A much simpler, although limited, approach is due to Lord Rayleigh. This will be outlined in the next section.

2.4.3 Rayleigh theory

For the Rayleigh approach to be valid, the particles must be much smaller than the light wavelength, so that all of the radiation from one particle will be in phase. In practice, this means that the particle diameter must be less

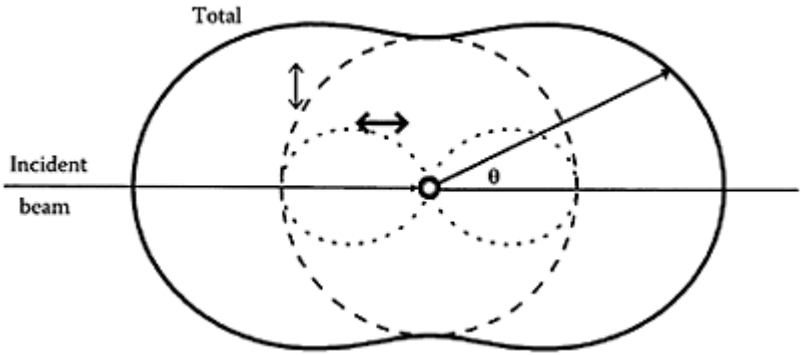

Figure 2.14 Angular distribution of scattered light according to the Rayleigh approximation, Equation (2.38), showing the total scattering *(full line)* and the components polarized in vertical and horizontal directions *(dashed and dotted lines)*.

than 10% of the wavelength (i.e., well below 100 nm for visible light). If the incident beam is unpolarized and of intensity I_0, then the scattered intensity at an angle θ to the beam is given by the following:

$$\frac{I_\theta}{I_0} = \frac{1}{r^2}\left[\frac{\pi^4 d^6}{2\lambda^4}\left(\frac{m^2-1}{m^2+2}\right)^2\left(1+\cos^2\theta\right)\right] = \frac{R_\theta}{r^2} \qquad (2.38)$$

Here, d is the diameter of the (spherical) particle, m is the relative refractive index (i.e., the refractive index of the particle relative to that of the medium), and r is the distance of the detector from the sample (see Figure 2.12). (The scattered light intensity follows the inverse square law; hence the dependence on $1/r^2$.) The term in square brackets is known as the *Rayleigh ratio*, R_θ.

In general, scattered light is polarized to some extent. In the $(1+\cos^2\theta)$ term in the Rayleigh equation, the 1 represents the vertically polarized component of the scattered light and $\cos^2\theta$ represents the horizontal component. (The terms "vertical" and "horizontal" are relative to the plane defined by the incident beam and the detector.) Thus the vertically polarized light has equal intensity at all scattering angles, whereas the

intensity of the horizontal component depends on the scattering angle, being maximum at 0 degrees and 180 degrees and zero at 90 degrees. This is illustrated in Figure 2.14, where the angular intensity distribution of scattered light is shown. The scattering pattern from very small particles is symmetric, in that equal amounts of light are scattered in the forward and backward directions.

It is also possible to calculate the *total* amount of light scattered by a particle by integrating the angular distribution over all directions. This can be used to calculate a scattering coefficient, given by the following:

$$Q = \frac{8}{3}\alpha^4\left(\frac{m^2-1}{m^2+2}\right)^2 \tag{2.39}$$

where α is a dimensionless size parameter, given by the following:

$$\alpha = \frac{\pi d}{\lambda} \tag{2.40}$$

For the Rayleigh theory to apply, α cannot exceed about 0.3. So, for typical values of relative refractive index f or particles in water (*m* up to about 1.4), Equation (2.39) gives very low values of Q—below about 10^{-3}. This means that the Rayleigh theory is only suitable for particles that effectively scatter less than 1/1000 of the light incident on them.

With the Rayleigh expression for the scattering coefficient, we can calculate the specific turbidity of a monodisperse suspension from Equation (2.37):

$$\frac{\tau}{\phi} = \frac{4\pi^4\alpha^3}{\lambda}\left(\frac{m^2-1}{m^2+2}\right)^2 \tag{2.41}$$

There are several important features of the Rayleigh results:

- The scattering intensity at a given angle depends on the *sixth* power of particle size, which means that scattering becomes very weak for smaller particles. For a fixed volume concentration, the turbidity depends on the *cube* of particle size. These results have serious implications for the detection and sizing of very small particles by optical methods. Suspensions of very small particles, such as colloidal silica, can have low turbidities, even at high solids concentrations.
- For a given mass (and hence total volume) of particles, increase of particle size (e.g., by aggregation) should lead to a substantial increase in scattered light, despite the decrease in particle number concentration. Equation (2.41) shows that specific turbidity increases as the cube of particle size. For colloidal particles, aggregation is often made visibly apparent by an increased turbidity
- Scattering depends inversely on the *fourth* power of the light wavelength and so becomes much stronger for shorter wavelengths. With incident white light, it is the blue end of the spectrum that is scattered most. This accounts for the blue color of the sky, red sunsets, and similar phenomena.

As already mentioned, expressions based on the Rayleigh theory are only applicable to very small particles (less than 0.1 μm for visible light).

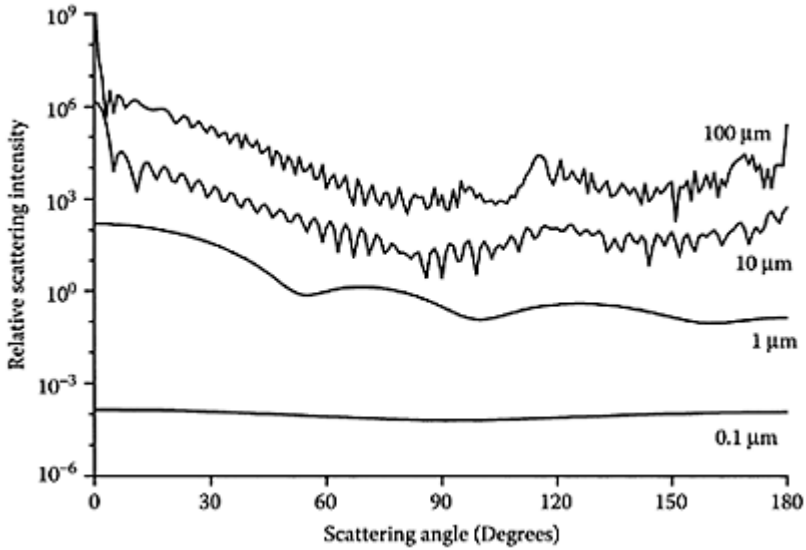

Figure 2.15 Scattered light intensity for different particle sizes (diameters shown on curves) as a function of scattering angle. The particles are assumed to have a relative refractive index, m=1.20, and the light wavelength is 650 nm. The computations are based on the Mie theory for the total (unpolarized) scattered light.

For larger particles these equations give results that are seriously in error and other approaches have to be used. The Mie theory gives exact solutions for spherical particles, and there are several other approximations available.

2.4.4 Mie theory

A complete and rigorous solution to the problem of light scattering by spheres, usually known as the Mie theory, has been available since early in the 20th century. Without going into details of the theory, we shall give some numeric results showing the essential features of light scattering by larger particles.

An important aspect is that the angular distribution of scattered light becomes highly asymmetric, as particles become larger, with a greater fraction of light being scattered in the forward direction. This is illustrated in Figure 2.15 for particles ranging in diameter from 0.1 µm (around the limit of applicability of the Rayleigh theory) to 100 µm. The

computations assume a wavelength of 650 nm and a relative refractive index, m=1.20. The results are plotted as relative scattered intensity versus scattering angle, and for simplicity only the total scattered intensity is shown. However, as with all scattered light, there are significant differences between the horizontally and vertically polarized components, especially at angles of around 90 degrees.

We can see from Figure 2.15 that the scattered light intensity varies by many orders of magnitude over the size range of 0.1–100 μm (note the logarithmic scale). At the low end of this range the intensity depends only

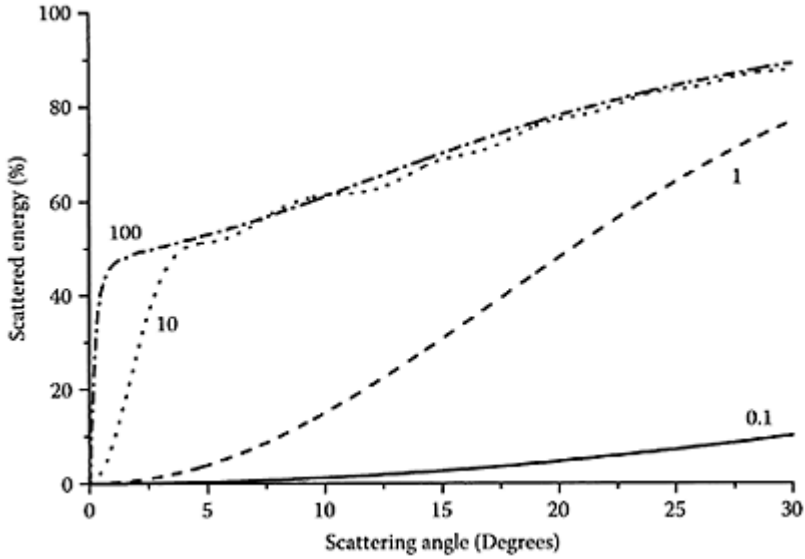

Figure 2.16 Proportion of total scattered light energy as a function of scattering angle. Same conditions as for Figure 2.15.

slightly on scattering angle, as expected from the Rayleigh theory (see Figure 2.14). For the larger particles, the scattering pattern is complex and the intensity at low angles is very much greater than at higher angles.

It is also possible to calculate the fraction of the total scattered energy within a given angle of the incident beam. Results of such calculations, for the same conditions as in Figure 2.15, are plotted in Figure 2.16, which shows that, for the larger particles, more than half of the scattered energy is found within forward angles of only a few degrees (only about 2 degrees for the 100-μm particles). This has important consequences for practical measurements of transmitted light because a significant amount of scattered light may reach the detector, giving an apparent turbidity lower than the actual value.

Scattering coefficients, Q, (see Section 2.4.2)) can also be calculated from the Mie theory for spherical particles of any size. Because Q also depends on light wavelength, it

is convenient to give results in terms of the dimensionless size parameter a, eq (2.40), because this includes both size and wavelength. The other important parameter is the relative refractive index, *m*. Scattering coefficients calculated for two values of m and over a range of α values are shown in Figure 2.17 (this also includes Q values from the anomalous diffraction approximation; see Section 2.4.5). For the m=1.2 case, Q values pass through a series of decreasing maxima at regular intervals. At values of a larger than those shown the scattering coefficient becomes approximately constant at Q=2. For the lower refractive index, similar behavior is found but over a more extended range of α values.

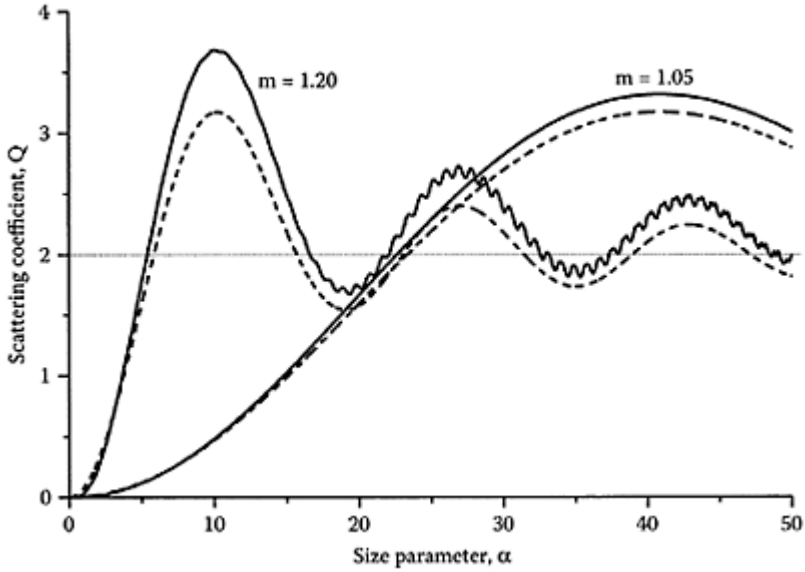

Figure 2.17 Scattering coefficient for two different values of refractive index, m, as a function of the dimensionless size parameter,. Full lines: Computations from the Mie theory. Dashed lines: Results from the anomalous diffraction approximation, Equation (2.42).

The fact that Q approaches a limiting value of 2 for large particles is worthy of comment. It means that a large particle removes from the beam *twice* the amount of light incident on it. This is often called the *extinction paradox*. For large particles essentially all of the incident light energy is scattered or absorbed, which would give Q=1. However, there is also the effect known as *diffraction*, which gives a characteristic angular pattern. Diffraction is a well-known effect in classical optics, and for a sphere the diffraction

pattern is just the same as a circular disk (or hole) of the same diameter. The total amount of diffracted light is also equal to the incident light, so that the total scattering efficiency is Q=2. Note that the full Mie theory implicitly includes diffraction, although, for large particles, scattering at small angles can be assumed to result entirely from diffraction and this forms the basis of a method to determine particle size (see Section 2.5.4).

From the scattering coefficient, it is possible to calculate specific turbidity from Equation (2.37); results are shown in Figure 2.18. Because the specific turbidity is defined in terms of absolute particle size, the results are plotted versus particle diameter, rather than size parameter a. For this reason, the results only apply to the chosen wavelength, 650 nm (this is typical of practical turbidity measurements). It is clear that specific turbidity depends greatly on particle size. For very small particles, the specific turbidity is low, but it increases steeply with increasing particle size. The first maximum occurs at a size that depends on refractive index. For *m*=1.2 (typical of, say,

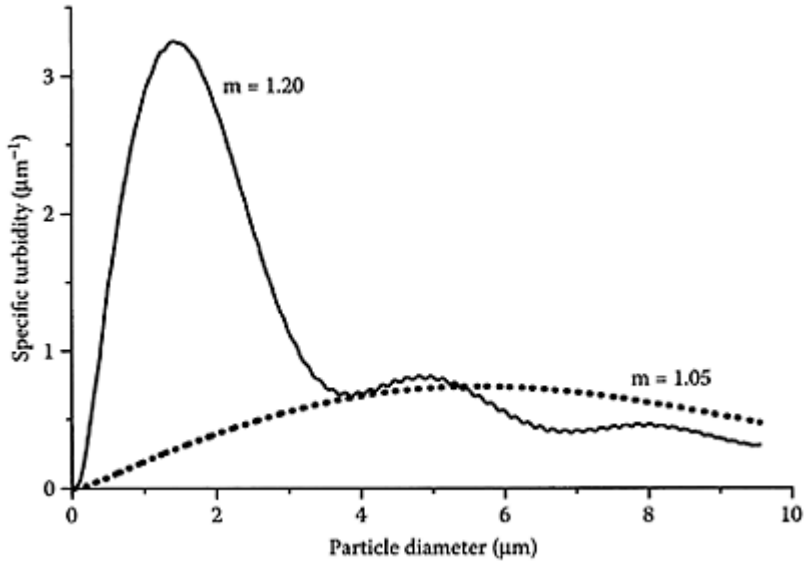

Figure 2.18 Specific turbidity for two values of refractive index, as a function of particle diameter. Wavelength=650 nm.

clay particles), the maximum is found at a particle diameter of about 1.4 μm. For lower refractive indices, the maximum is shifted to larger particle sizes. It is also noteworthy that the maximum turbidity for *m*=1.2 is about four times greater than that for the lower refractive index. After the first maximum, there is a general decrease in specific turbidity.

A more common practical method of measuring turbidity is by light scattering at some angle to the incident beam, usually 90 degrees *(nephelometry).* The scattered intensity at 90 degrees shows an upward trend with increasing particle size. For a suspension,

however, it is more useful to think in terms of scattered intensity *per unit volume of particles* (analogous to the specific turbidity). Such calculated values, based on the Mie theory, for a wavelength of 650 nm and the same refractive index values as before are shown in Figure 2.19. Both show a broadly similar form, but, for the lower refractive index, *m*=1.05, the results are at least 10 times lower than for *m* =1.20. As with specific turbidity in Figure 2.18 there is a steep increase as particle size increases, followed by a maximum. In the 90 degrees scatter case, this first maximum is sharp and occurs at a diameter of about 0.3 μm, much lower than for specific turbidity. This is followed by other maxima and minima, but the general trend is downward.

The results in Figures 2.18 and 2.19 show that conventional turbidity measurements, either by transmission or 90-degree scattering, are highly dependent on particle size and become less sensitive for large particles. This has important implications f or monitoring particles in water. Turbidity measurements are not well-suited to sensitive detection of particles larger than a few μm in size.

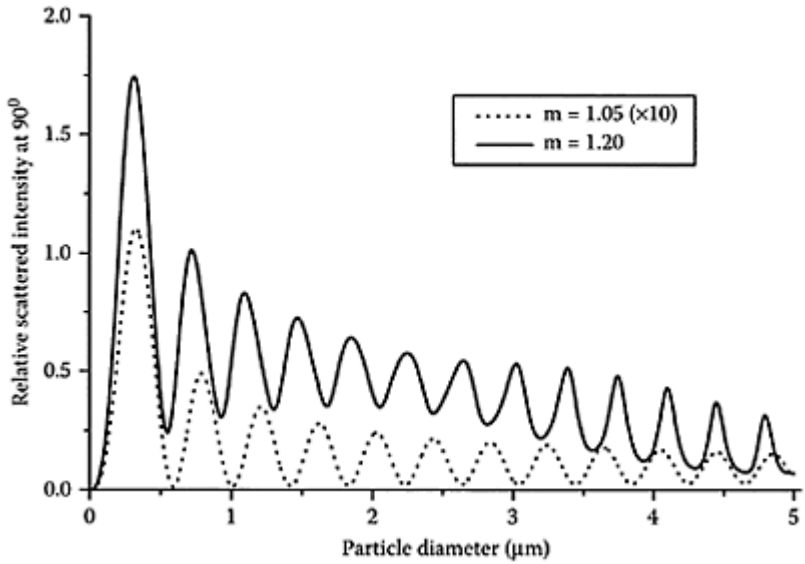

Figure 2.19 Relative scattered light intensity *per unit volume of particles*, as a function of particle diameter, for two values of refractive index and a wavelength of 650 nm. Note that the values for m=1.05 are multiplied by a factor 10 for better visibility.

Although calculations based on the Mie theory are straightforward, using modern computers and readily available software, there is still interest in approximate methods for certain applications. Two such approximations will be briefly discussed.

2.4.5 Anomalous diffraction

For particles that are appreciably larger than the light wavelength and with a fairly low refractive index, it can be assumed that light rays pass through a sphere with little deviation. This allows a greatly simplified treatment, which leads to *the anomalous diffraction approximation*. The most useful result from this approach is a convenient approximation for the scattering coefficient:

$$Q = 2 - \left(\frac{4}{\rho}\right)\sin\rho + \left(\frac{4}{\rho^2}\right)(1-\cos\rho) \qquad (2.42)$$

where $\rho=2\alpha(m-1)$.

Equation (2.42) is a reasonably good approximation to exact Mie results provided that the parameter ρ is not too small and *m* is low. Results from Equation (2.42) were plotted in Figure 2.17 for two values of refractive index. For *m*=1.20, the anomalous diffraction result is about 14% too low at the first maximum, but this error becomes less for larger sizes. For *m*=1.05, the error at the first maximum is only about 4%. However, for small particles (values of $\rho \ll 1$) Equation (2.42) gives values of Q that are much too high.

2.4.6 Rayleigh-Gans-Debye scattering

An expression that is useful in some cases is generally known as the *Rayleigh-Gans-Debye (RGD)* approximation. The physical basis of this approach is that a particle (of any shape) is assumed to consist of elements that behave as independent Rayleigh scatterers. This is appropriate when the parameter ρ (see earlier) is very low, which implies small particle size and low refractive index. Although the RGD result can be used for particles that are too large for the Rayleigh approximation to apply, the condition $\rho \ll 1$ means that the scattering coefficient Q must be very small.

The RGD expression is a modified form of the Rayleigh result, Equation (2.38):

$$\frac{I_\theta}{I_0} = \frac{1}{r^2} R_\theta P(\theta) \qquad (2.43)$$

where R_θ is the Rayleigh ratio (Equation [2.38]) and $P(\theta)$ is known as a *form factor*.

The form factor is a correction to the Rayleigh expression, which accounts for effects resulting from the size and shape of the particle. This can be evaluated for many simple geometric shapes and, importantly, for aggregates of particles in terms of the mass distribution. For homogeneous spheres, diameter *d*, the form factor is as follows:

$$P(\theta) = \left[\frac{3(\sin u - u\cos u)}{u^3}\right]^2 \qquad (2.44)$$

where $u=qd/2$ and *q* is the *scattering vector*, which depends on the scattering angle and the light wavelength:

$$q = \frac{4\pi}{\lambda}\sin\frac{\theta}{2} \qquad (2.45)$$

The most important application of the RGD approximation is to the case of particle aggregates. This aspect will be considered in Chapter 5.

2.5 Measurement of particle size

There are many approaches to the problem of determining particle size and a large number of commercial instruments, based on various techniques. There is no "universal" method applicable over the whole size range, and choices have to be made, depending on the nature of the suspension. We shall give here only a very broad survey of the more common methods, without going into technical detail.

One important general point has already been mentioned at the beginning of this chapter. This is the question of particle shape. In determining particle size it is very convenient to report the result in terms of a single parameter, such as an "equivalent diameter," although, for nonspherical particles, this gives no indication of the true shape. Mostly, we just have to accept that reported particle sizes are subject to this limitation.

2.5.1 Direct methods (microscopy)

One of the oldest and still the most direct method of determining particle size is by microscopic observation. A sample of suspension can be viewed at an appropriate magnification and individual particles can be sized, using a suitable scale or by an automated image analysis method.

Optical microscopy is limited by the wavelength of visible light, so that particles smaller than around 1 μm are difficult to resolve. In fact, reliable particle sizes cannot be derived for particles smaller than about 5 μm. It is possible to use dark-field illumination *(ultramicroscope),* by which particles are seen as points of light against a dark background. This is essentially a light scattering method and allows small particles to be resolved, although reliable sizing is difficult.

A more recent development, from around 1990, is the *confocal laser scanning microscope,* in which a small sample of suspension is illuminated by a laser beam. The beam scans the sample point by point and layer by layer and the scattered light is monitored by a detector. In this way, a three-dimensional image can be built up with a much higher resolution than a conventional light microscope.

Electron microscopy, because of the much lower effective wavelength, is capable of far higher resolution than optical microscopy, and particle sizes of a few nanometers can be reliably determined. Sample preparation is more elaborate and may introduce artifacts.

A major advantage of direct microscopic observation is that particle shapes are clearly apparent. However, it is still common to report equivalent diameters, and several conventional methods may be used, some of which are illustrated in Figure 2.20.

- The *projected area diameter* is simply the diameter of a circle having the same area as the projected area of the particle.
- The *Feret diameter* is defined as the perpendicular distance between two parallel tangents, in a fixed direction, to the projected outline of the particle.
- The *Martin diameter* is the length of a chord, in a fixed direction, which divides the projected area of the particle into two equal parts.

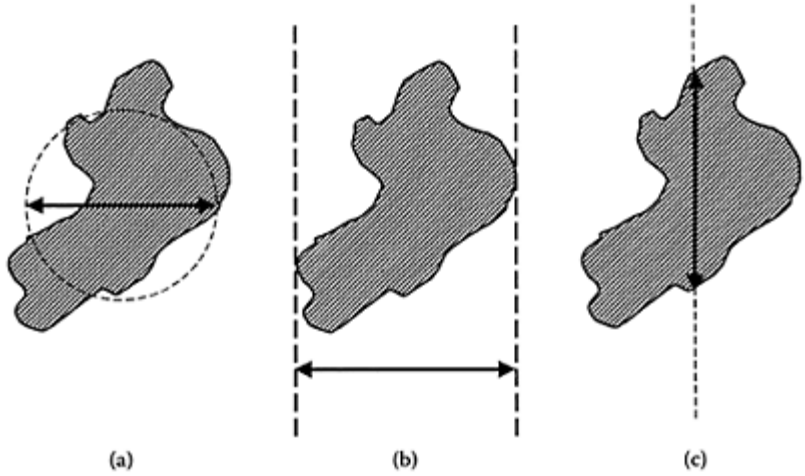

Figure 2.20 Different ways of deriving a "diameter" from microscope images of nonspherical particles. (a) Projected area (equivalent circle) diameter; (b) Feret diameter; (c) Martin diameter. (See text.)

It is important to note that, in the case of the Feret and Martin diameters, the procedure is to choose a fixed direction for the tangents and chords for all particles measured. This takes care of the fact that particles on the microscope stage lie in random directions and gives an appropriate average value.

It is possible to give more detailed information on irregular particles, but some simplification is still needed. For instance, elongated particles may be approximated by ellipsoids and the major and minor axes of the projected ellipse can be reported.

2.5.2 Particle counting and sizing

Automated particle counting can be achieved by allowing particles to pass singly through a zone in which their presence can be detected by a suitable sensor (see Figure 2.21). Particles passing through the zone produce a response from the detector and can be counted as a series of pulses. If the sensor response depends on particle size, then the pulse height is size-dependent and this provides a means of discriminating between particles of different sizes.

It is important to ensure that particles do not pass simultaneously through the sensing zone; otherwise two or more particles would produce just one pulse and would be counted as one (larger) particle. This is known as the *coincidence* effect, and, although statistical correction for this effect is possible, the most satisfactory procedure is to ensure that the sample is sufficiently dilute to avoid the problem. For many suspensions of

interest, high degrees of dilution are necessary, which may cause some change in the particle size distribution, especially for aggregating particles.

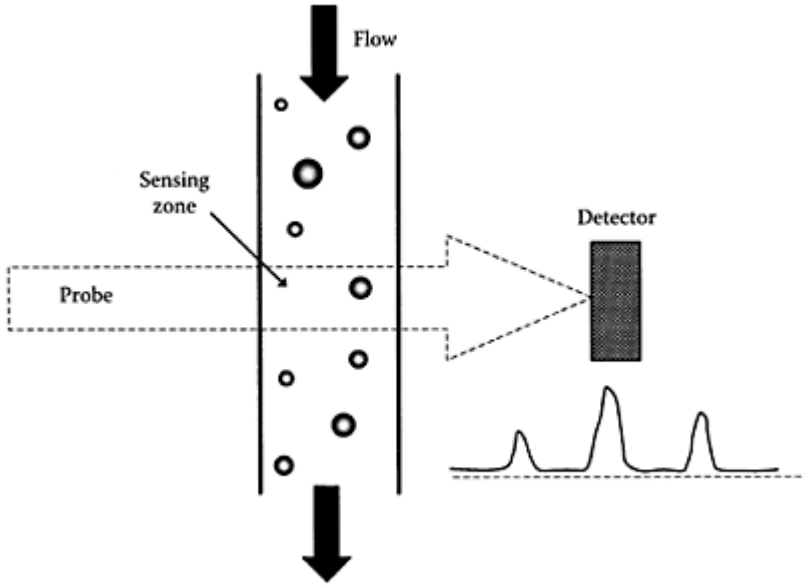

Figure 2.21 Schematic diagram of a particle counting and sizing technique. Particles should pass singly through the sensing zone, generating pulses with a suitable probe/ detector system.

There are two commonly used detection techniques in particle counting: electrical and optical, giving *electrical sensing zone (electrozone)* and *optical sensing zone (light scattering)* counters. The two types have their own advantages and disadvantages.

Electrozone counters are based on a principle developed in the late 1940s by Wallace Coulter and made available commercially as the *Coulter counter*. The basic concept (often called the *Coulter principle*) is that particles in an electrolyte solution passing through an orifice cause a momentary change in electrical resistance and hence a voltage pulse if the current between two electrodes is maintained constant. The electrodes are located on either side of the orifice, and a particle passing through displaces a volume of electrolyte equal to the particle volume. Most particles can be assumed to have infinite resistance, in comparison to the electrolyte solution, and so the voltage pulse produced is proportional to the particle volume.

The Coulter technique has been used widely, with thousands of references in the literature. It is capable of high-speed counting (5000 or more particles per second) and can resolve particles only slightly different in size. The most common use of this technique has been for rapid counting of blood cells.

A unique feature of the electrozone method, in comparison with other (especially optical) techniques, is that it is virtually independent of the shape or composition of the particles. A particle passing through the orifice gives a pulse that depends only on the volume of electrolyte displaced and hence on the volume of the particle. For an aggregate, the pulse height is proportional to the total volume of the constituent particles, without the included fluid (because this is electrolyte solution it does not contribute to the extra resistance).

An often-quoted difficulty in using the electrozone technique to study aggregation phenomena is the possibility of aggregate breakage during passage through the orifice, where there may be very high shear rates. This question is not straightforward, however, because breakage of an aggregate gives fragments of the same total volume and the size registered by the counter should be the same as that of the original aggregate. Thus, breakage within the orifice should give one pulse, corresponding to the aggregate size. Breakage in the elongational flow field just ahead of the orifice could give several smaller pulses.

Another problem is that high degrees of dilution are often necessary to avoid coincidence effects. The dilution fluid has to be a particle-free salt solution (usually 2% NaCl), which would affect the colloidal stability of particles (see Chapter 4). However, at the low particle concentration needed to avoid coincidence (usually less than around 10^6 particles per mL) aggregation would be very slow (see Chapter 5) and should not be significant in the short time needed for a measurement.

The electrozone technique cannot conveniently be used over a very wide range of particle size. To detect particles reliably, an orifice is needed with a diameter not more than about 50 times the primary particle size. Particles smaller than about 0.5 μm are difficult to monitor, which is a serious limitation for colloids. Particles larger than about 40% of the orifice diameter also present difficulties because orifice blockage can become a serious problem. Thus, for a 50-μm orifice, particles in the size range of 1–20 μm could be measured. For a wider size range 2 or more orifice sizes have to be used, which is inconvenient.

Particle counting and sizing can also be achieved using *light scattering* techniques. In this case, particles are made to pass singly through a focused light beam (usually a laser beam) and either the transmitted light or scattered light intensity is measured. Each particle passing through the beam causes a reduction in transmitted light or an increase in the scattered light (see Figure 2.12). The decrease in transmitted light is caused by scattering at all angles and should be much greater than the increase in scattered light at any particular angle. Nevertheless, the scattering technique is the more sensitive and can be used for smaller particles than the transmission (or *light blockage*) method. The reason is that detecting particles by light transmission depends on reliably measuring a small difference between large quantities (the transmitted light intensities with and without a particle in the beam). In the case of scattered light, the value in the absence of a particle should be essentially zero, so that even a very small increase can be detected. Lower particle size limits are usually about 1 μm for light blockage counters and around 0.2 μm for light scattering instruments. In both cases, the size of particles can be derived using standard light-scattering (Mie) theory (see Section 2.4.4). Because light scattering depends on the refractive index of particles, optical counting and sizing results are difficult to interpret for suspensions containing particles of different composition, as in

natural waters. The usual practice is to calibrate the instrument with monodisperse particles of various sizes (such as unif orm latex suspensions), so that the pulse height can be related to particle size. However, this correlation only applies to particles having the same shape and refractive index as the spherical latex particles, which is by no means the case for all aquatic particles. This difficulty is often ignored in practice, but it means that particle size information from optical particle counters, either by light "blockage" or scattering, must be interpreted with some caution.

A newer method of counting and sizing particles is *focused beam reflectance measurement (FBRM)*. A laser beam is projected into a suspension of particles through a window and focused to a small spot close to the window. This spot is caused to rotate at high speed (around 2 m/s). As a particle passes by the window, the focused beam will pass one edge of the particle and some of the light is backscattered until the beam passes the opposite edge. By analyzing the backscattered light signal, and knowing the speed of the rotating light spot, it is possible to determine a *chord length* between two points on the edge of the particle. With sufficient particles, at random orientations to the path of the rotating spot, an average chord length can be derived, characterizing the average particle size. It is also possible to generate a chord length distribution, which is related to the particle size distribution. However, for nonspherical particles there will be a distribution of chord lengths simply as a result of the different orientations of particles in the beam. This technique is able to measure particles over a wide size range (0.5 μm–2.5 mm) and is applicable to much more concentrated suspensions than the methods described earlier. Also, like the electrozone method, no knowledge of the particle properties is needed.

A big advantage of all of these particle counting and sizing techniques is that information on the particle size distribution is derived more or less directly, with no assumption about the form of the distribution needed.

2.5.3 Static light scattering

We saw earlier in the chapter that light scattering is greatly dependent on the size of particles and so, in principle, it provides a powerful experimental technique for the determination of particle size.

Even simple measurements of turbidity by transmission should be useful. Figure 2.18 shows that the specific turbidity depends greatly on particle size, but the shape of the curve makes it impossible to derive a size unambiguously from a single measurement. A given value of specific turbidity could correspond to at least two particle sizes. By carrying out measurements at two or more wavelengths it may be possible to get around this problem, but the result would still only be an average particle size, with no inf ormation on the size distribution.

Measurement of scattered light as a function of scattering angle is more promising as a method of determining particle size. Except in the Rayleigh scattering region (see 2.4.3.), the angular scattering pattern is dependent on particle size, but, again, it is difficult to derive information on the size distribution.

The usual way around the problem of particle size distribution is to assume a convenient mathematical form (such as the log-normal distribution) and to fit parameters of the distribution to the experimental scattering data.

Another problem with light-scattering methods is that information on the particle refractive index is needed, which effectively restricts the method to suspensions of the same material. This limitation does not apply to large particles, where a simple diffraction method may be applicable.

2.5.4 Fraunhofer diffraction

For particle sizes very much greater than the light wavelength, scattering can be treated as a problem in geometric optics, as mentioned earlier. A large spherical particle in a light beam can be treated as a circular disc with the same diameter. At the edge of the disc, light is diffracted and gives a characteristic pattern of light and dark rings at a plane far from the particle. This has become known as *Fraunhofer diffraction*. These bands represent maxima and minima in the intensity of the diffracted light and their positions are dependent only on the light wavelength and the particle diameter, not on other particle properties. Although diffraction by particles has been known since the early 19th century and was used to study the size of blood cells in 1918, it only achieved widespread use from about 1980.

For particles larger than about 10 μm and for optical wavelengths the bands occur at low angles. Nevertheless, using laser illumination; good-quality optics, including a Fourier transform lens; and an array of concentric detectors, it is possible to derive detailed information. Several commercial instruments using diffraction methods are available, and these find very wide application for routine particle size analysis. Since the simple Fraunhofer theory becomes unsuitable for particles not much larger than the light wavelength, commercial software usually includes computations based on the Mie theory. This means that information on the nature of the particles (such as refractive index) is then needed.

For heterodisperse suspensions it is necessary to invert the diffraction data to give a particle size distribution. The only feasible way of doing this is to assume a form of the size distribution, such as log-normal and to derive the appropriate parameters by an iterative procedure. Using such methods, available instruments are claimed to give particle size over a very wide range (usually 1–1000 μm).

2.5.5 Dynamic light scattering

The basis of dynamic light scattering is that light scattered from a moving particle will have a slightly different frequency from the incident light, just as in the well-known *Doppler effect*. In colloidal dispersions, random brownian motion of the particles causes scattered light to vary randomly in frequency and interference between light scattered from different particles causes random fluctuations in the light intensity measured by a stationary detector. A characteristic “speckle pattern” is visible when a suspension of brownian particles is illuminated by a laser beam. Analysis of this effect involves autocorrelation of the scattered light intensity, measured as a train of pulses from a photomultiplier. The nature of the technique is such that it may be called *photon correlation spectroscopy (PCS)* or *quasi-elastic light scattering (QELS)* as well as dynamic light scattering. Practical implementation was made possible when coherent laser light and dedicated correlators became available. Commercial instruments were

introduced in the early 1980s and are widely used for sizing of particles in the submicron range. (Because the effect is dependent on brownian motion, a practical upper limit of particle size is about 3 μm.) From the autocorrelation function of the photon pulse train, it is possible to derive the particle diffusion coefficient D.

Because the diffusion coefficient of spherical particles is related to their size by the Stokes-Einstein relation, Equation (2.27), dynamic light scattering provides a measure of particle size, at least for monodisperse, spherical particles. For nonspherical particles, the size given is that of an equivalent sphere, with the same diffusion coefficient.

For polydisperse samples, derivation of the particle size distribution is mathematically difficult. Commercial instruments usually give a mean size, together with a "polydispersity index," which gives an indication of the breadth of the size distribution.

2.5.6 Sedimentation methods

Because sedimentation rate is greatly dependent on particle size, at least for dilute suspensions, it provides a convenient method for size determination. Such techniques are long established and are used widely

Sedimentation can be studied by gravimetric techniques, such as by various forms of sedimentation balance or by some kind of pipet method. These are tedious, but easy to perform. A more convenient approach is to use a *photosedimentation* device, which combines gravitational settling with light transmission measurements. In some cases the light beam and detector can be moved up and down a settling column, or multiple beams and detectors can be used, so that the measurement time is reduced. Of course, to interpret the results in terms of particle size, information on the density and light-scattering properties of the particles must be available. For particles all of the same density, it is not too difficult to derive a size distribution.

X-rays have been used instead of light beams. In that case, particles are so much larger than the effective wavelength that their scattering cross-sections are directly proportional to the projected area (see Section 2.4.4.).

With gravitational settling, these techniques are restricted to particles of at least a few μm in size because smaller particles settle too slowly for convenience. This problem can be overcome by using a centrifuge technique, often combined with light or X-ray transmission analysis. In this way, the sedimentation technique can be extended well into the submicron size range.

All particle sizes derived by sedimentation techniques are equivalent Stokes diameters.

Further reading

Allen, T., *Powder Sampling and Particle Size Determination,* Elsevier Science, New York, 2003

Kerker, M., *The Scattering of Light and Other Electromagnetic Radiation,* Academic Press, New York, 1969.

Stanley-Wood, N.G. and Lines, R.W., *Particle Size Analysis,* Royal Society of Chemistry, Cambridge, 1992.

van de Hulst, H.C., *Light Scattering by Small Particles,* Dover Publications, New York, 1981

Vogel, S., *Life in Moving Fluids,* Princeton University Press, 1996.

chapter three
Surface charge

3.1 Origin of surface charge

A particle in contact with an aqueous solution is likely to acquire a surface charge for various reasons. The most common reason is that the surface has chemical groups that can ionize in the presence of water to leave a residual charge on the surface, which can be either positive or negative. This and some other important mechanisms will be discussed briefly in the next sections.

3.1.1 Dissolution of constituent ions

Many crystalline solids, such as calcium carbonate, have limited solubility in water and their particles can acquire charge because one or other of the constituent ions has a greater tendency to "escape" into the aqueous phase. In the early colloid literature, there are many studies of silver halides, especially *silver iodide*, AgI, and this provides a useful example, although it is not especially relevant to particles in natural waters.

Silver iodide has very low solubility in water, and its solubility constant at room temperature is about 10^{-16} (mol L^{-1})^2. In pure water, the concentration of silver and iodide ions would each be about 10^{-8} mol/L, and under these conditions the particles are negatively charged. Essentially, silver ions have a greater tendency than iodide to enter the aqueous phase, leaving a net negative charge on the crystal. This tendency arises from differences in binding of ions to the crystal lattice and in hydration of ions in aqueous solution. By changing the relative concentrations of Ag^+ and I^-, (e.g., by adding NaI or $AgNO_3$ to the solution) it is possible to change the surface charge. A condition can be found at which the net surface charge is zero. This is the *point of zero charge* (pzc), which is a very important concept in colloid science. By increasing the silver ion concentration in solution to about 3.2×10^{-6} mol/L, the preferential tendency for Ag^+ to migrate into water is just balanced by its excess concentration in solution, so the net surface charge becomes zero. Under these conditions the iodide concentration has to be about 3×10^{-11} mol/L to maintain the correct value of the solubility constant.

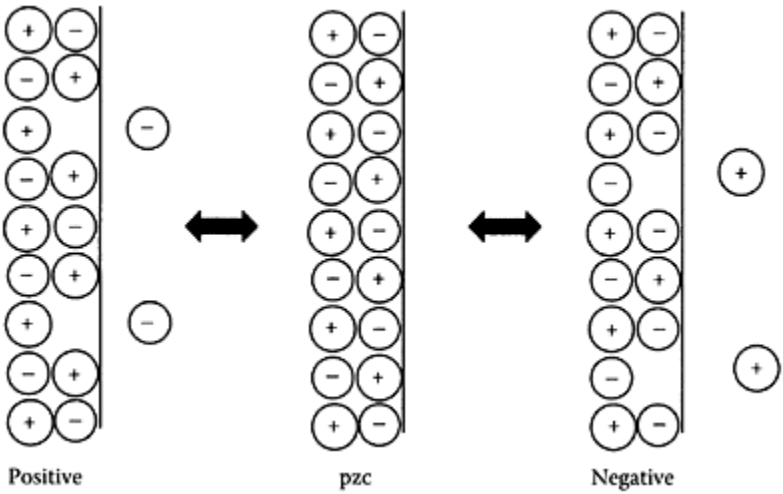

Figure 3.1 Development of surface charge by an ionic solid of low solubility, whose constituent ions are potential determining. The point of zero charge (pzc) occurs at a certain concentration of these ions in solution. If the cation concentration is greater than that at the pzc, then the surface charge is positive and vice versa.

The schematic diagram in Figure 3.1 illustrates the concept of pzc for a crystalline solid.

In such cases, the constituent ions are known as *potential determining ions* (pdi). Straightforward thermodynamic reasoning leads to a relationship between the concentration of pdi and the electric potential of the solid relative to the solution. The latter is effectively the *surface potential* and is given the symbol ψ_0. This is related to the concentration (strictly thermodynamic *activity*) of the potential determining ions by the *Nernst* equation:

$$\psi_0 = \text{constant} + \frac{RT}{z_i F} \ln c_i \qquad (3.1)$$

where R is the universal gas constant, T is the absolute temperature, z_i is the valence, c_i is the concentration of the potential-determining ion, and F is the Faraday constant.

It follows that the rate of change of surface potential with pdi concentration is as follows:

$$\frac{d\psi_0}{d(\log_{10} c_i)} = 2.303 \frac{RT}{z_i F} \qquad (3.2)$$

For a singly charged ion such as Ag^+, (z_i=1), Equation (3.2) implies that the surface potential changes by about 59 mV for a 10-fold change in pdi concentration. The Nernst equation is used in a number of areas and is directly applicable to ion-selective electrodes.

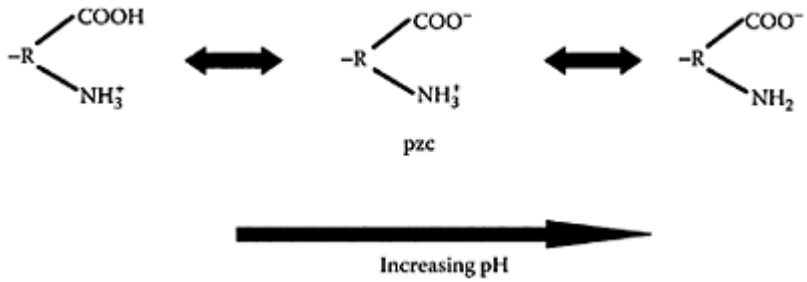

Figure 3.2 Showing the ionization of surface carboxylic and amine groups. In this case, the point of zero charge (pzc) occurs at a particular pH value.

Similar considerations apply to calcium carbonate, where the pdis are Ca^{2+} and CO_3^{2-}. However, this case is complicated by equilibria between the carbonate ion and HCO_3^-, which gives an apparent pH dependence of the surface potential.

3.1.2 Surface ionization

Many surfaces have acidic or basic groups at which protons (H^+) can either be released or acquired, depending on the pH of the solution. Biological surfaces provide an important example. These usually have proteins as part of the surface structure, which have carboxylic (COOH) and amine (NH_2) groups. These are weakly acidic and basic groups, respectively, which ionize in the manner shown in Figure 3.2.

At low pH the carboxylic groups are not dissociated and hence uncharged, whereas the amine groups are protonated and have a positive charge. At high pH the carboxylic groups dissociate to give a negative charge and the amine groups lose their proton and are uncharged. So, such surfaces are positively charged at low pH and negatively charged at high pH.

There is a characteristic pH value at which the number of negatively charged surface groups just balances the number of positive groups. By analogy with the case in Section 3.1.1, this is also called the pzc, although H^+ is not a constituent ion of the particle and so it is not strictly a potential-determining ion as previously defined. The pzc depends on the number and type of surface groups and their respective ionization equilibria. For many biological surfaces (e.g., of bacteria and algae), the pzc is in the region of pH 4–5, so that in most natural waters such particles are negatively charged. For a related reason, many inorganic particles in the aquatic environment are negatively charged by virtue of an adsorbed layer of natural organic matter.

Another very important example of surface ionization is that of metal oxides, such as alumina Al_2O_3, ferric oxide, Fe_2O_3, titania, TiO_2, and many others. In water, the surfaces of oxide particles become hydroxylated to give surface groups such as AlOH, which are amphoteric (i.e., they can ionize to give either positive or negative charge). In highly simplified form, the ion-

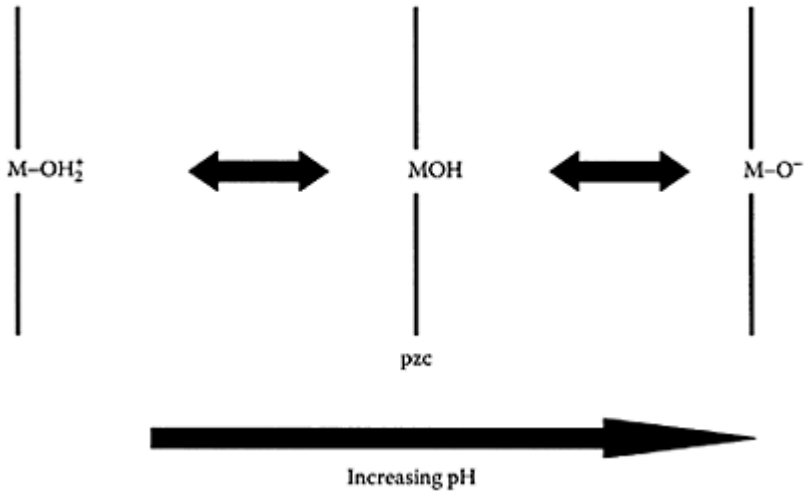

Figure 3.3 Ionization of metal hydroxide (MOH) groups at an oxide surface. The point of zero charge (pzc) occurs at a certain pH value.

ization of surface groups metal hydroxide can be represented as shown in Figure 3.3.

Again, the surface is positively charged at low pH and negatively charged at high pH, with a characteristic pzc. For oxides, pzc values depend on the acid-base properties of the metal and vary over a wide range. Also, the precise value depends on the crystalline form of the oxide, origin, preparation details, and the presence of impurities, so it is difficult to quote definitive pzc values. However, the following give a rough indication for some important oxides:

Oxide:	SiO_2	TiO_2	Fe_2O_3	Al_2O_3	MgO
pzc:	2	6	8	9	12

Acidic materials, such as silica, have a great tendency to lose protons and are negatively charged over most of the pH range. By contrast, the basic oxide MgO acquires protons readily and is positively charged up to about pH 12. Intermediate cases, such as ferric oxide, have pzc values around neutral pH, so that the surface charge may be positive or negative in natural waters.

Although H^+ (and OH^-) are not potential determining ions in the strict thermodynamic sense, it is often found that the surface potentials of oxides and similar materials show Nernst-like dependence on pH, especially in the region of the pzc.

Some surfaces have only one type of ionic group, as for latex particles with carboxylic or sulfate groups. In these cases, the degree of ionization and hence the surface charge may be dependent on pH, but there is no pzc. The surface charge becomes less negative as the pH is reduced, but there is no charge reversal.

3.1.3 Isomorphous substitution

Some materials have an "inherent" excess charge as a result of isomorphous substitution. The best-known examples are clay minerals, such as kaolinite, which has an alternating two-layer structure with tetrahedral silica and octahedral alumina layers. In the silica layers some Si^{4+} ions may be replaced by Al^{3+}, and in the alumina layers Al^{3+} may be replaced by Mg^{2+}. In both cases the lattice is left with a residual negative charge, which must be balanced by an appropriate number of "compensating cations." These are usually fairly large ions such as Ca^{2+}, which cannot be accommodated in the lattice structure, so that these ions are mobile and can diffuse into solution when the clay is immersed in water. This gives rise to the well-known *cation exchange* properties of clays.

At low pH the edges of kaolin particles can acquire positive charge, as for oxides, whereas the faces have negative charge because of isomorphous substitution. For this reason the particles can aggregate in an edge-to-face (or "house of cards") structure.

3.1.4 Specific adsorption of ions

Even when a surface has no ionizable groups or inherent charge, it is still possible for it to become charged by adsorption of certain ions from solution. Adsorption of ions on a neutral surf ace is *specific* in the sense that there must be some favorable interaction other than electrostatic attraction. Ions adsorbing on an oppositely charged surface may do so for purely electrostatic reasons, and this could not be responsible for the development of surface charge. Conversely, adsorption of ions on surfaces with the same sign of charge must involve some favorable "chemical" interaction to overcome electrostatic repulsion.

A well-known example is the adsorption of surfactant ions to give a surface charge. Typical surfactants have a hydrocarbon "tail," which is hydrophobic, and a hydrophilic head group, which may be ionized. The hydrophobic part can minimize contact with water by adsorbing on a hydrophobic surface, as shown schematically in Figure 3.4. In this way, with ionic surfactants the surface can acquire charge. Surfactants can be used to stabilize oil droplets, air bubbles, and many types of solid particle for this reason.

Many metal ions can adsorb on surfaces in a "specific" manner, by forming coordinate bonds with groups on the surface. Good examples are metal ions at oxide surfaces, where surface complex formation can give strong adsorption and charge reversal. In such cases, the adsorbing metal ion must lose some of its water of hydration and is said to form an "inner sphere complex" with the surface group. If a fully hydrated ion adsorbs on an oppositely charged surface, it forms an "outer sphere complex" and is held only by electrostatic attraction. Hydrolysis of metal ions such as aluminum and iron can lead to

stronger specific adsorption on surfaces, and this is an important factor in the action of hydrolyzing metal coagulants (see Chapter 6).

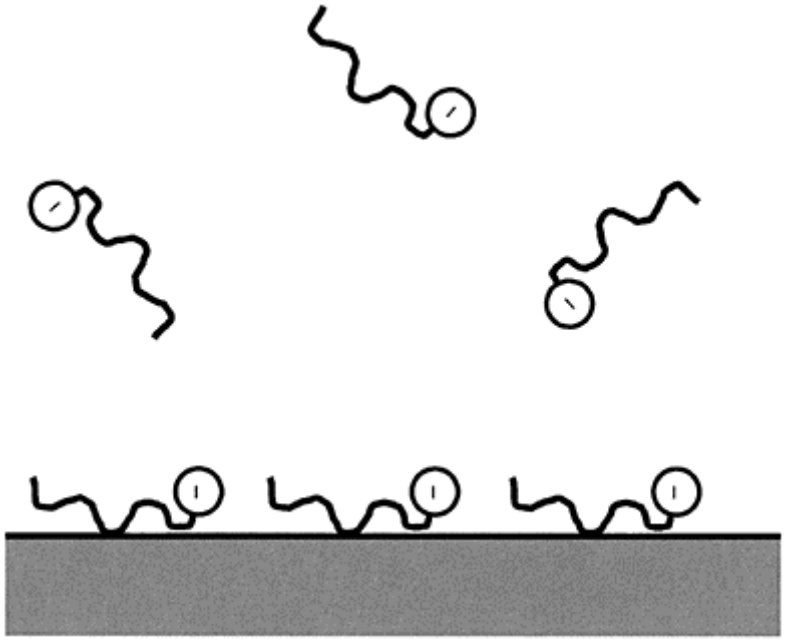

Figure 3.4 Schematic diagram showing the adsorption of an anionic surfactant on a hydrophobic surface.

Even simple ions can show specific differences in adsorption by virtue of differences in hydration. Anions tend to be less strongly hydrated than cations and so can approach closer to a surface, giving an apparent negative surface charge. This effect may partly explain the tendency of otherwise neutral surfaces to become negatively charged in aqueous salt solutions. The negative charge of air bubbles in water may be an example of this effect.

3.2 The electrical double layer

Whatever the origin of surface charge, a charged surface in contact with a solution of ions will lead to a characteristic distribution of ions in solution. If the surface is charged, then there must be a corresponding excess of oppositely charged ions *(counterions)* in solution to maintain electrical neutrality. The combined system of surface charge and the excess charge in solution is known as the *electrical double layer*. This is an extremely important topic in colloid science that has been studied in great detail over many decades, leading to

theoretical models of varying complexity. We shall only consider a rather simple model, which nevertheless conveys the essential properties of the double layer.

3.2.1 The double layer at a flat interface

It is convenient initially to consider a flat charged surface in contact with an aqueous salt solution. The counterions are subject to two opposing tendencies:

- Electrostatic attraction to the charged surface
- The randomizing effect of thermal motion

The balance between these effects determines the distribution of charge and electric potential in solution. The first serious attempts to model the double layer were made independently by Gouy and Chapman in the early years of the 20th century. The model is based on a number of simplifying assumptions:

- There is an infinite, flat, impenetrable surface.
- Ions in solution are point charges able to approach right up to the charged surface.
- The solvent (water) is a uniform medium with properties that are not dependent on distance from the surface.

This approach leads to a prediction of the way in which the electric potential in solution varies with distance from the charged surface. For fairly low values of surface potential, the potential in solution falls exponentially with distance from the surface. The main difficulty with the Gouy-Chapman model is the assumption of ions as point charges. In fact, real ions have a significant size, especially if they are hydrated, and this limits the effective distance of closest approach to a charged surface. Allowing for the finite size of ions gives a region close to the surface that is inaccessible to counterion charge. This has become known as the *Stern layer*, after Otto Stern, who first introduced the ion size correction into double layer theory in 1924. The Stern layer contains a certain proportion of the counterion charge, and the remaining counterions are distributed within the diffuse part of the double layer or simply the *diffuse layer*.

A conceptual picture of the Stern-Gouy-Chapman model of the electrical double layer at a flat interface is given in Figure 3.5. This shows the variation of electric potential from the surface, where its value is ψ_0, to a distance far into the solution, where the potential is taken as zero. Across the Stern layer, the potential falls rather rapidly, to a value $\psi\delta$ (the *Stern potential*) at a distance δ from the surface (this is at the boundary of the Stern layer, known as the *Stern plane*). Usually, δ is of the order of the radius of a hydrated ion or around 0.3 nm. Although this distance is very small, the Stern layer can have a great influence on double layer properties. From the Stern plane into the solution, through the diffuse layer, the potential varies in an approximately exponential manner, according to the following:

$$\psi=\psi_\delta \exp(-\kappa x) \tag{3.3}$$

where x is the distance from the Stern plane and κ is a parameter that depends on the concentration of salts in the water. This approximation is strictly only valid for fairly low values of the Stern potential, but this is not often a serious limitation in practice.

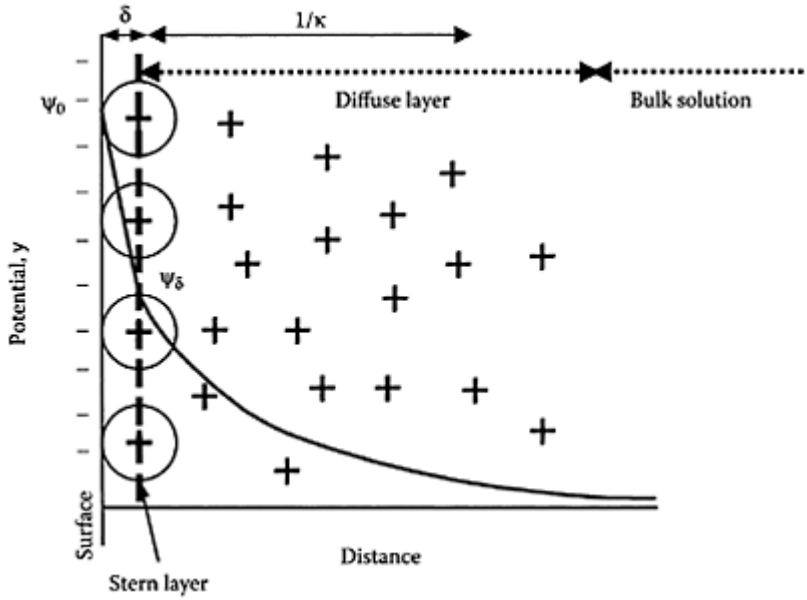

Figure 3.5 The Stern-Gouy-Chapman model of the electrical double layer adjacent to a negatively charged surface (see text).

It is worth pointing out that Figure 3.5 only shows the excess counterions in the diffuse layer. Generally there will be various dissolved salts in water and hence a range of cations and anions. In fact, because of electrostatic repulsion, there will also be a *deficit* of co-ions (anions in this case) close to the charged surface. Far away from the charged surface the concentrations of cations and anions will have values appropriate to the bulk solution and their charges will exactly balance. All of the surface charge is compensated by excess counterions (and deficit of co-ions) in the double layer region. The system as a whole (charged surface and solution) is electrically neutral. For simplicity, only counterions (cations) are shown in Figure 3.5.

The parameter κ plays a large part in the interaction of charged particles in water and is known as the *Debye-Hückel parameter*. To calculate the value of κ we need to know the concentration, c_i, and charge (valence), z_i, of all significant ions in solution, together with certain physical quantities, such as the universal gas constant, R, the absolute temperature, T, Faraday's constant, F, and the permittivity of the solution, ε, which is equal to the relative permittivity ε_r (or dielectric constant) multiplied by the permittivity of free space, ε_0. The value of κ is then given by the following:

$$\kappa^2 = \frac{1000F^2}{\varepsilon RT}\sum(c_i z_i^2) \qquad (3.4)$$

The summation is taken over all ions present in solution and is related to the *ionic strength, I*, which is defined as follows:

Table 3.1 Typical values of the Debye length 1/κ

Solution	1/κ (nm)
Pure (deionized) water	960
10^{-4} M NaCl	30
10^{-4}M $CaCl_2$	18
10^{-3}M $MgSO_4$	5
River Thames	4
Sea water	0.4

$$I = \frac{1}{2}\sum(c_i z_i^2) \qquad (3.5)$$

Note that Equation (3.4) is given in rationalized (SI) units. In some older texts a different version may be found. The factor 1000 is included because ion concentrations, c_i, are conventionally given in mol/L rather than mol/m^3.

The parameter κ has dimensions of 1/length (m^{-1}), and 1/κ is sometimes known as the *Debye length* or the "thickness" of the double layer. It can be seen from Equation (3.3) that, when x=1/κ, the potential in the diffuse layer has fallen to 1/e of the Stern potential. The Debye length is essentially a scaling parameter, which determines the extent of the diffuse layer and hence the range over which electrical interaction operates between particles. It can be seen from Equation (3.4) that, as the ion concentration and/or valence increases, κ increases and hence the Debye length decreases. This effect is sometimes referred to as *double layer compression* and is highly relevant to the stability of colloidal particles (see Chapter 4).

If numeric values appropriate for water at 25°C are used, the parameter κ is related to ionic strength by the following:

$$\kappa = 3.29\sqrt{I}\ (\mathrm{nm}^{-1}) \qquad (3.6)$$

For typical salt solutions and natural waters, values of the Debye length 1/κ can range from less than 1 nm to around 100 nm or more. For completely deionized water at 25 °C, the concentrations of H^+ and OH^- are each 10^{-7} M and the Debye length is about 1000 nm (or 1 μm). Some examples are given in Table 3.1.

3.2.2 Charge and potential distribution in the double layer

We have seen that there is a characteristic decrease of electric potential away from a charged surface. It is also useful to know how the counterion *charge* is distributed.

The surface is assumed to have a charge density of σ_0 (C/m^2). If there is no specific adsorption (see later) then all of the surface charge must be balanced by excess charge in the diffuse layer, σ_d, so that $\sigma_0 = -\sigma_d$. It can be shown from basic electrostatics that the total charge per unit area in the diffuse layer is equal to the gradient of the potential at the inner boundary of the diffuse layer (i.e., at the Stern plane). It follows that the charge in the diffuse layer is directly proportional to the Stern potential, if the latter is fairly small:

$$\sigma_d = -\varepsilon\kappa\psi_\delta \tag{3.7}$$

(The minus sign is necessary because the diffuse layer charge must be opposite in sign to the surface charge; for a negative surface charge the diffuse layer charge is positive).

Equation (3.7) is like the charge-potential relationship for a parallel plate capacitor, with a capacitance $C_d = \varepsilon\kappa$. It follows that the effective distance between the "plates" is $1/\kappa$, which is the Debye length. This is one reason why the Debye length can be regarded as the effective "thickness" of the double layer. In fact, the combination of the Stern layer and diffuse layer can be regarded as two parallel plate capacitors in series. If the capacitance of the Stern layer is C_S, the total capacitance C_T is given by the standard formula:

$$\frac{1}{C_T} = \frac{1}{C_S} + \frac{1}{C_d} \tag{3.8}$$

The total potential drop over both capacitors is ψ_0, which is divided into potentials across the Stern layer, $\psi_0 - \psi_\delta$, and the diffuse layer, ψ_δ (see Figure 3.6). The Stern layer capacitance can be regarded as fixed for a given system. (It depends on the Stern layer thickness, δ, and the effective permittivity of water close to the surface, which is usually much less than that of ordinary bulk water.) However, the capacitance of the diffuse layer depends on the ionic strength, via the Debye-Hückel parameter, κ. Because κ increases with ionic strength, it follows that the diffuse layer capacitance also increases. This means that, with increasing salt concentration, a smaller proportion of the total potential drop occurs across the diffuse layer and hence a larger proportion occurs over the Stern layer. In other words, the Stern potential, ψ_δ, decreases with increasing ionic strength. This is illustrated schematically in Figure 3.6.

It can be seen from Figure 3.6 that increasing ionic strength has two important effects on double-layer properties:

- A decrease in the Stern layer potential
- A decrease in the "thickness" (compression) of the diffuse layer

Both of these effects occur with *any* added salt, and those that act only in this way are known as *indifferent electrolytes*. However, because of the strong effect of ion valence on

the parameter κ, it turns out that highly charged ions will be more effective than singly charged ions.

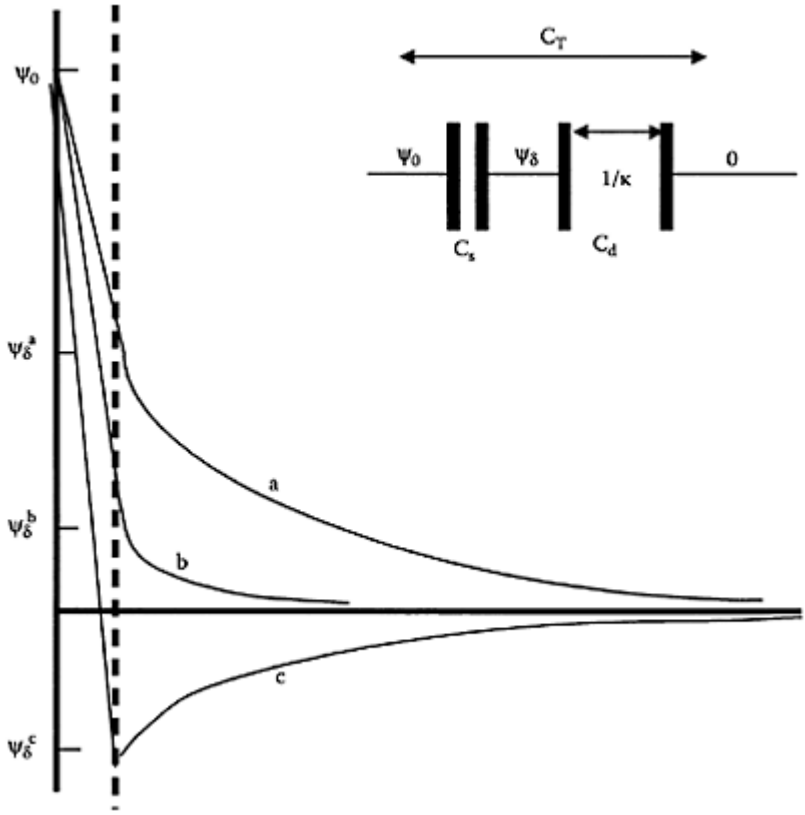

Figure 3.6 Potential distribution in the double layer for the following: (a) low concentration of "indifferent" electrolytes; (b) high concentration of indifferent electrolytes; and (c) low concentration of a salt with specifically adsorbing cations, giving charge reversal. The inset shows the representation of the double layer as capacitors in series.

In some cases the added salt may contain ions that are *specifically adsorbed* on the surface (see Section 3.1.4). If these have an opposite charge to the surface (counterions), then it is possible for the surface charge to be reversed, as shown in Figure 3.6c. For

strongly adsorbed counterions, quite low concentrations can cause charge reversal, so that there may be a fairly small effect on the extent of the diffuse layer. Specific adsorption, with surface charge reduction, is a very important practical method for destabilization of charged particles (see Chapter 4).

3.2.3 Spherical particles

So far, we have been entirely concerned with flat charged surfaces, but these are not appropriate models for many real particles in water. It is not possible to deal with particles of arbitrary shape, so we shall briefly consider electrical double layers associated with charged spheres.

A complete theoretical treatment of spherical double layers is difficult; we shall only deal with a simple approximation. This is for the case in which

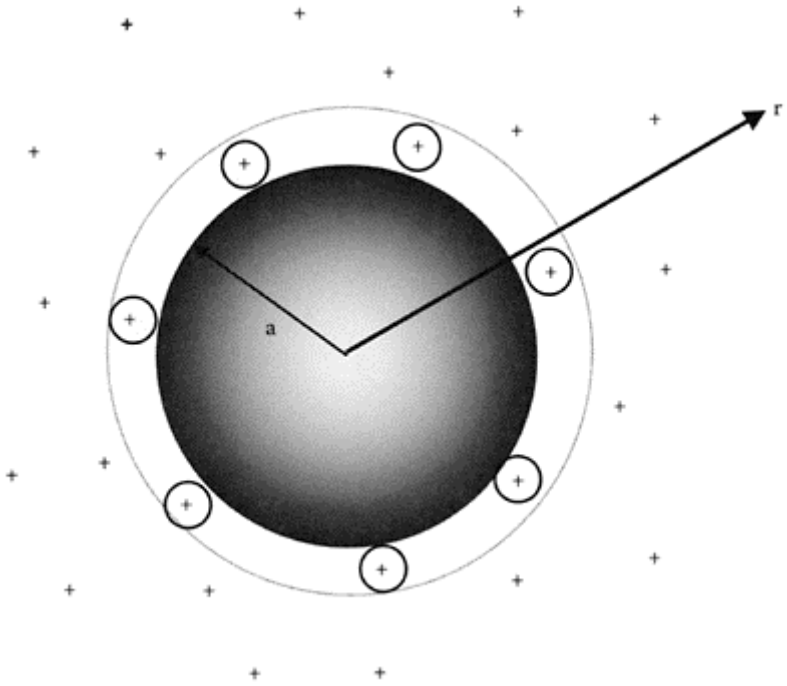

Figure 3.7 The electrical double layer for a negatively charged spherical particle.

the Stern potential is quite low and the variation of potential with radial distance r from the center of the sphere, with radius a (see Figure 3.7), is also given by a simple exponential form:

$$\psi = \frac{\psi_\delta a}{r}\exp\left[\kappa(a-r)\right] \tag{3.9}$$

The surface charge density on the sphere is given by the following:

$$\sigma_0 = \frac{\varepsilon\psi_\delta}{a}(1+\kappa a) \tag{3.10}$$

When the double layer is relatively thin compared to the particle size (or $\kappa a >> 1$), these expressions become equivalent to the corresponding equations for a flat surface, Equations (3.3) and (3.7).

3.3 Electrokinetic phenomena

When there is relative movement between a charged interface and the adjacent solution, some of the charge in the diffuse layer moves with the solution, giving rise to *electrokinetic* effects.

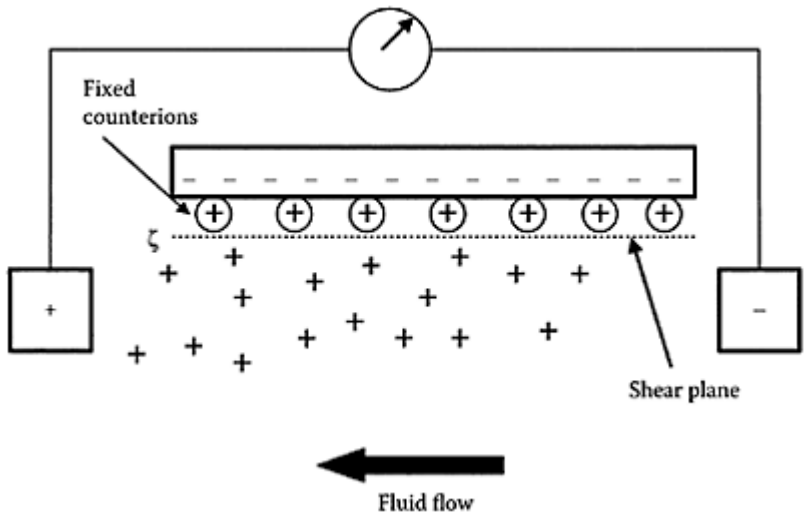

Figure 3.8 Development of a streaming potential as a result of fluid flow past a charged surface. The movement of counterions in the mobile part of the double layer, outside the shear plane, causes the left-hand electrode to acquire a positive potential relative to the right-hand electrode.

The potential at the shear plane is the *zeta potential*, ζ, which can be derived from such measurements.

3.3.1 The plane of shear and the zeta potential

A fundamental concept in electrokinetics is that there is a *plane of shear*, which separates the fixed and mobile charge. Between this plane and the surface, it is assumed that the liquid and any ions present in this region are effectively "fixed" to the surface. Outside the plane of shear, the ions in the diffuse part of the double layer are free to move with the liquid. This is shown schematically in Figure 3.8, where a negatively charged flat surface, and hence positive counterions, are assumed.

If the solution moves in the direction shown, then some of the counterion charge will move from right to left. This would lead to the development of an electric potential difference such that the left-hand (downstream) electrode shown would become positive relative to the upstream electrode. This is one of several possible electrokinetic effects and is known as *streaming potential*. This and other effects will be discussed below.

Electrokinetic measurements can be used to derive the electric potential at the plane of shear, which is universally known as the *zeta potential* and given the symbol ζ. The precise location of the shear plane and hence the relation between the zeta potential and the double layer structure has been a source of much debate. For most practical purposes it can be assumed that the plane of shear is close to (perhaps just outside) the Stern plane (Figure 3.5), so that the zeta potential is approximately equal to the Stern potential ψ_δ There is no satisfactory method of testing this assumption because there is no independent technique for measuring the Stern potential. However, the zeta potential has attained enormous importance in colloid science, not least because it can be measured fairly easily. Electrical interaction between charged particles is greatly dependent on the zeta potential (Chapter 4).

The various electrokinetic phenomena can be observed either by measuring an electric potential (or current) when relative motion is induced between a charged surface and a liquid or by applying an electric field and observing the relative motion caused. The known effects are as follows:

1. *Electro-osmosis*, which is the flow of liquid through a porous plug (or tube) of material under the influence of an applied electric field.
2. *Particle electrophoresis*, which is the movement of charged particles in an electric field.
3. *Streaming potential*, which arises when liquid is forced to flow through a porous plug (or tube) of material. A streaming *current* can also be measured.
4. *Sedimentation potential*, which can be observed when charged particles fall through a liquid.

Of these, the most useful for characterizing particle charge is the electrophoresis technique, and several commercial instruments are available, such as the *Zeta Meter*, *Zetasizer*, and others. Instruments based on the streaming current technique are also used, especially for optimizing coagulant dosages in water treatment *(Streaming Current Detector, SCD)*. We shall restrict our attention here to particle electrophoresis.

3.3.2 Electrophoresis and electrophoretic mobility

When an electric field is applied to a suspension of charged particles, the particles move toward one of the electrodes. Negatively charged particles move toward the positive electrode and vice versa. The velocity of the particles is directly proportional to the applied field strength, and the ratio of these two quantities is known as the *electrophoretic mobility* (EM). The usual units for expressing EM are $\mu m\ s^{-1}/V\ cm^{-1}$, and on this scale most values lie in the range of −5 to +5. More formally the SI units for EM should be $m^2s^{-1}V^{-1}$ ($1\ \mu m\ s^{-}1/V\ cm^{-1}=10^{-8}\ m^2s^{-1}V^{-1}$.)

The measurement of electrophoretic mobility is straightforward in principle and simply involves measuring the velocity of particles in a field of known strength. Older techniques (still in widespread use) use a specially designed cell and direct microscopic observation of the moving particles. The Zeta Meter and the Rank Brothers Mark 2 apparatus are examples of the direct technique. More recently a method that uses *Laser Doppler Velocimetry* has been adopted. Essentially crossed laser beams produce interference fringes of known spacing. Particles moving through the fringes scatter light with a fluctuating intensity, the frequency of which is directly related to the particle velocity.

A significant complication associated with electrophoretic mobility arises from the fact that the walls of the observation cell are usually charged. This causes an electro-osmotic flow, which is superimposed on the particle velocity. The usual way around this problem is to carry out measurements at a *stationary level*, where the electro-osmotic flow is zero. (In a closed cell, the flow along the walls is accompanied by a return flow around the center and there are well-defined levels where these two flows exactly cancel.) At the stationary levels the observed particle velocity is the true electrophoretic velocity. In some cases, the inner walls of the electrophoresis cell are coated with a gel, which is essentially uncharged, so that there is little or no electro-osmotic flow. This avoids the need to locate the stationary levels.

A quite different experimental approach is based on electroacoustics. When a suspension of particles is subjected to high-frequency sound (ultrasonic) waves, the particles are caused to vibrate at the same frequency. If the particles are charged and hence surrounded by a diffuse layer of counterions, the particle vibration causes some charge separation and generates a measurable alternating electric field. The charge separation arises essentially because the ions are much more mobile than the particles and can respond more rapidly to the ultrasonic waves. The particles lag behind the counterions, creating a small, fluctuating electric dipole. The cumulative effect of many such dipoles gives a *colloid vibration potential*, the magnitude of which is proportional to the electrophoretic mobility of the particles. For basically the same reason, application of an alternating electric field to a suspension of charged particles generates sound waves, an effect known as *electrokinetic sonic amplitude* (ESA).

These effects are exploited in commercial instruments to determine the zeta potential of particles in concentrated suspensions, for which conventional electrophoretic mobility techniques and laser Doppler methods are not applicable. However, electroacoustic methods cannot be used with dilute suspensions (typically less than about 1% solids).

Having determined the electrophoretic mobility, how is the zeta potential derived? In the general case, this can be complicated, but there are two simple approximations that apply to limiting situations.

One of these is the case in which the particles are very small, so that the diffuse layer is of much greater extent. Mathematically this condition can be stated as $\kappa a << 1$, where κ is the Debye-Hückel parameter and *a* is the particle radius. In this case, the zeta potential is related to the EM value, *U*, by the *Hückel equation:*

$$U = \frac{2\varepsilon\zeta}{3\mu} \tag{3.11}$$

where μ is the viscosity of the liquid.

Most particles in water are too large for this approximation to be applicable. The Hückel equation is more suited to ions in solution and polyelectrolytes.

When the diffuse layer is very thin compared to the particle size $(\kappa a >> 1)$ the *Smoluchowski equation* may be used, which differs from Equation (3.11) by a factor of 1.5:

$$U = \frac{\varepsilon\zeta}{\mu} \tag{3.12}$$

This is a much more useful expression in practice. It can be applied to particles of around 1-μm diameter or larger in aqueous salt solutions of ionic strength of about 1 mM or greater. Furthermore, the Smoluchowski equation applies to particles of any shape, provided that the diffuse layer is very thin compared to any radius of curvature. Inserting values of permittivity and viscosity for water at 25°C in Equation (3.12), a very simple relation between zeta potential and mobility is found:

$$\zeta(\text{mV}) = 12.8U(\mu\text{m s}^{-1} / \text{V cm}^{-1}) \tag{3.13}$$

So, the zeta potential (in mV) is derived simply by multiplying the mobility (in standard units) by a factor of about 13.

For intermediate particle sizes, where the previously detailed approximations do not apply, several complications arise and it is a much less straightforward matter to derive zeta potential from electrophoretic mobility. For this reason, it is common practice to quote experimental results in terms of mobility, rather than derived zeta potentials.

Further reading

Hunter, R.J., *Zeta Potential in Colloid Science,* Academic Press, London, 1981.
Koopal, L.K., Adsorption of ions and surfactants, in *Coagulation and Flocculation,* Dobias, B., Ed., Marcel Dekker, New York, 1993.
Stumm, W., *Chemistry of the Solid-Water Interface,* Wiley-Interscience, New York, 1992.

chapter four
Colloid interactions and colloid stability

4.1 Colloid interactions—general concepts

Between particles in water there are various kinds of interaction, depending on the properties, especially surface properties, of the particles. These interactions can give forces of attraction or repulsion. If attractive forces dominate, then the particles will stick together on contact and form clusters or *aggregates*. When particles repel each other they are kept apart and prevented from aggregating. In the latter case, the particles are said to be *stable*, whereas when aggregates can form, the particles are *unstable* or *destabilized*. Because these concepts are mainly relevant to particles in the colloidal size range, the subject is known as *colloid stability* and the interactions are known as *colloid interactions*. These topics are dealt with in this chapter.

4.1.1 Importance of particle size

Before dealing with the different types of colloid interactions, it is worth pointing out some important general features. The first point is that colloid interactions are usually of rather short range—usually much less than the particle size. Thus, they do not come into play until particles are nearly in contact and so do not have much influence on the transport of particles, which is still governed by mechanisms discussed in Chapter 2 (i.e., diffusion, sedimentation, and convection). When particles *do* approach very close, then colloid interactions are crucial in determining whether attachment occurs.

The other important feature is the dependence of the interactions on particle size. As we shall see, in most cases, the strength of interaction is roughly proportional to the *first power* of the particle size. There are other important forces acting on particles, which were discussed in Chapter 2 (i.e., fluid drag and gravitational attraction). Fluid drag is proportional to the projected area of the particle and thus roughly to the *square* of the particle size. The gravitational force is proportional to the mass of the particle and hence to the *cube* of particle size.

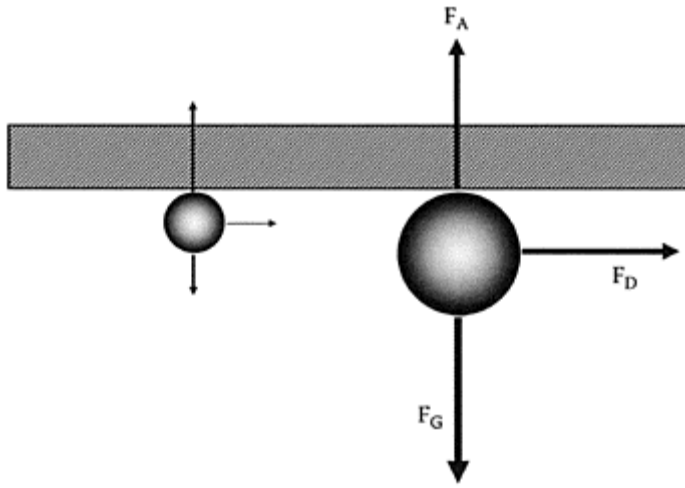

Figure 4.1 Forces on a sphere close to a flat plate. F_A: Attractive colloid force (such as van der Waals attraction); F_D: Fluid drag; F_G: Gravitational attraction. For the smaller particle *(left)* the colloid force is largest, but for the larger particle the other forces are relatively more significant.

These different size dependences are of enormous significance because they mean that colloid interactions become less important as particle size increases. This is illustrated schematically in Figure 4.1. Here two spherical particles are in contact with a flat wall and are subject to three different forces:

- An attractive force, F_A, holding the particles to the wall
- Fluid drag, F_D, caused by flow parallel to the surface
- Gravitational attraction, F_G, acting vertically downward, opposite to F_A.

The two particles have diameters that differ by a factor of two, and the magnitude of the forces is indicated by the lengths of the arrows. For the smaller particle, the attractive force is greater than the gravitational force and hence the particle remains attached. However, for the larger particle, although F_A is doubled, F_G is greater by a factor of 8 than for the smaller one. This means that gravity would be sufficient to detach the larger particle. The drag force is increased by a factor of 4, and this may also play a part in detaching the larger particle. This simple example provides an explanation for the common observation that colloid interactions are much more significant for smaller particles and that larger particles can more easily be detached by fluid drag or other external forces, such as gravity. This is the main reason why the effects are known as *colloid* interactions.

4.1.2 Force and potential energy

The various types of colloid interaction give rise to *forces* between particles, which can be directly measured in some cases. However, it is often convenient to think in terms of a *potential energy* of interacting particles. These two concepts can be related simply by considering the work involved in bringing the particles from a large (effectively infinite) distance, where the interaction is negligible, to a given separation *h*. This gives the energy of interaction. If the interaction force at a separation distance *x* is *P(x)*, then the work done in moving through a distance δx is $P(x)\delta x$. Thus, the total work done in bringing the particles to a separation *h*, or the potential energy of interaction, *V*, is as follows:

$$V = \int_{h}^{\infty} P(x)dx \qquad (4.1)$$

Conventionally the sign of the force is positive for repulsion and negative for attraction, and the same applies to the energy of interaction.

It is usually easier to derive the interaction force, and then the corresponding energy can be calculated using Equation (4.1).

4.1.3 Geometry of interacting systems

A common method of approaching the problem of interaction between particles is to first derive the interaction between parallel flat plates as a function of separation distance. Some aquatic particles *are* platelike in character (e.g., clays), but in many cases we need to consider the interaction of particles that are roughly spherical in shape. It is possible to derive approximate expressions for interaction of spherical particles (and other shapes) by a method developed by Deryagin in 1934. We shall consider only two cases: the interaction of unequal spheres and between a sphere and a flat surface. Both of these are relevant to many common problems involving colloid interactions and are illustrated in Figure 4.2, together with the parallel plate case.

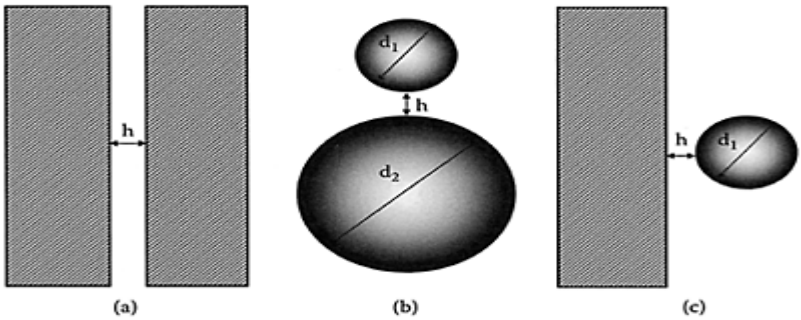

Figure 4.2 Showing interactions between (a) parallel flat plates, (b) unequal spheres, and (c) a sphere and a

plate. In all cases the separation distance is *h*.

The Deryagin approach makes the assumption that the interaction between spheres can be treated as the sum of interactions between concentric parallel rings, instead of the actual sphere surfaces. This approximation is only valid when the separation distance is much less than the sphere diameter. However, because colloid interactions are usually of small range, this is not often a serious limitation. If the *energy* of interaction per unit area of parallel plates, separated by a distance *h*, is *V(h)*, then the interaction *force* between unequal spheres, diameters d_1 and d_2, turns out to be simply:

$$P(h) = \frac{\pi d_1 d_2}{d_1 + d_2} V(h) \quad \text{(sphere–sphere)} \tag{4.2}$$

For the case of a sphere interacting with a flat plate, the force can be derived simply from the sphere-sphere case by letting one sphere become very large ($d_2=\infty$):

$P(h)=\pi d1V(h)$ (sphere-plate)

(4.3)

(This is just twice the force between two equal spheres of diameter d_1, at a distance *h*).

If the energy of interaction is needed, rather than the f orce, then Equation (4.1) can be used, with the appropriate force expression. It must be remembered that Equations (4.2) and (4.3) are only appropriate for very small separation distances $(h << d_1)$. They become inaccurate for larger distances.

4.1.4 Types of interaction

The following types of colloid interaction are important in practice and will be discussed in subsequent sections:

- van der Waals (usually attractive)
- Electrical double layer (either repulsive or attractive)
- Hydration effects (repulsive)
- Hydrophobic (attractive)
- Steric interaction of adsorbed layers (usually repulsive)
- Polymer bridging (attractive)

The first two of these interactions (van der Waals and electrical double layer) form the basis of a quantitative theory of colloid stability developed around 1940 by Deryagin and Landau and, independently, by Verwey and Overbeek. In recognition of these pioneers, the theory is now widely known as *DLVO theory*. The remaining interactions are not taken into account in the theory; these are sometimes called *non-DLVO forces*.

These interactions and their eff ects on colloid stability will be considered in the following sections.

4.2 van der Waals interaction

4.2.1 Intermolecular forces

Between all atoms and molecules there are attractive forces of various kinds, which J.D. van der Waals postulated in 1873 to account for the nonideal behavior of real gases. If the molecules are polar (i.e., with an uneven distribution of charge), then attraction between dipoles is important. When only one of the interacting molecules has a permanent dipole, then it can induce an opposite dipole in a nearby molecule, thus giving an attraction. Even when the atoms or molecules are nonpolar, the movement of electrons around nuclei give "fluctuating dipoles," which induce dipoles in other molecules and hence an attraction. From the standpoint of colloid stability this is the most important of the intermolecular interactions. It is a quantum-mechanical effect, first recognized by Fritz London in 1930. For this reason the resulting forces are sometimes known as *London-van der Waals forces*. However they are also known as *dispersion forces* because the fundamental electron oscillations involved are also responsible for the dispersion of light. (This term may be a source of some confusion—it does not refer to dispersions of particles.)

All of these interactions show the same distance dependence—the energy of attraction between molecules varies inversely as the *sixth* power of separation distance, *r:*

$$V(r) = -\frac{B}{r^6} \tag{4.4}$$

where *B* is a constant that depends on the properties of the interacting molecules (often known as the *London constant*) and the negative sign indicates an attraction.

The dependence on $1/r^6$ shows that the interaction falls off rapidly with increasing distance. However, between macroscopic objects, the attraction is of longer range and plays a vital part in the interaction of colloidal particles.

4.2.2 Interaction between macroscopic objects

All objects are assemblies of atoms and molecules, subject to the intermolecular interactions just discussed. In principle, the total interaction between two objects, of known geometric form, can be derived by adding up all of the individual intermolecular attractions. The summation is replaced by integration over the volumes of the interacting objects, and the result depends on the number of molecules per unit volume and the appropriate London constant, *B*. Such an approach was adopted by H.C. Hamaker in the 1930s; he showed that the resulting interactions could be appreciable at large separations. Some results are given below.

For *two parallel flat plates* separated by a distance *h*, the van der Waals energy of attraction per unit area is found to be the following:

$$V_A = -\frac{A_{12}}{12\pi h^2} \tag{4.5}$$

This expression is based on the assumption that the plates are "infinitely thick." (In practice this means that the thickness should be much greater than the separation distance.) The equation applies to the case where the two plates are composed of different materials, 1 and 2. The constant A_{12} is known as the *Hamaker constant*, which depends on the properties of the two materials. It is given by the following:

$$A_{12} = \pi^2 N_1 N_2 B_{12} \tag{4.6}$$

where N_1 and N_2 are the numbers of molecules per unit volume in the two materials and B_{12} is the London constant for the interaction of molecules 1 and 2.

Hamaker constants will be discussed further in the next section.

For the interaction of *unequal spheres*, the Hamaker expression is as follows:

$$V_A = -\frac{A_{12}}{12}\left[\frac{y}{x^2+xy+x} + \frac{y}{x^2+xy+x+y} + 2\ln\frac{x^2+xy+x}{x^2+xy+x+y}\right] \tag{4.7}$$

where $x=h/d_1$ and $y=d_2/d_1$.

For the sphere-flat plate case, the second sphere is assumed to be infinitely large ($y=\infty$), and the interaction energy becomes the following:

$$V_A = -\frac{A_{12}}{12}\left[\frac{1}{x} + \frac{1}{x+1} + 2\ln\frac{x}{x+1}\right] \tag{4.8}$$

These expressions can be considerably simplified if it is assumed that the separation distance is very small $(x << 1)$. This gives

$$V_A = -\frac{A_{12}}{12h}\frac{d_1 d_2}{(d_1+d_2)} \quad \text{(sphere–sphere)} \tag{4.9}$$

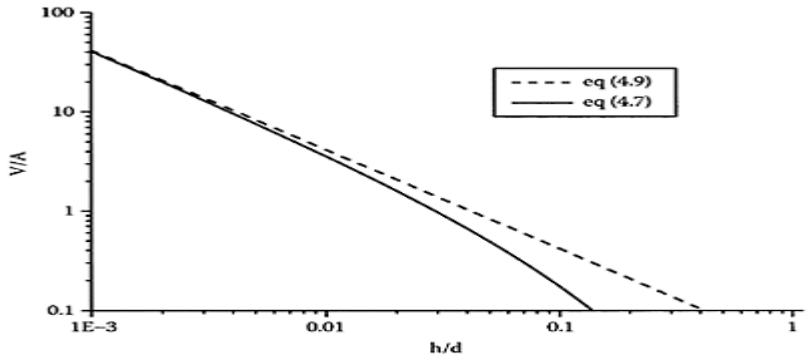

Figure 4.3 Comparison of van der Waals attraction between equal spheres, calculated from the complete Hamaker expression, Equation (4.7),

and the approximate, short-range expression, Equation (4.9).

$$V_A = -\frac{A_{12} d_1}{12h} \quad \text{(sphere-plate)} \tag{4.10}$$

These short-range expressions can also be derived by applying the Deryagin method, Equations (4.2) and (4.3), using the flat-plate energy expression, Equation (4.5). This gives expressions for the force, which, when integrated according to Equation (4.1), yield the same results as Equations (4.9) and (4.10). These are not good approximations when the separation distance exceeds a few percent of the particle diameter. Figure 4.3 shows calculated values of the interaction energy for two equal spheres as a function of dimensionless separation distance *h/d*. The energy is also expressed in dimensionless form, as a ratio with the Hamaker constant, *V/A*. Although not very accurate, the short-range expressions are adequate for many practical purposes.

It is clear that the van der Waals attraction between macroscopic objects has a different dependence on distance than that between molecules. The flat-plate energy depends inversely on the square of the separation distance, and for spheres at close approach there is a $1/d$ dependence. This means that the interaction falls much more slowly with increasing distance than the $1/r^6$ behavior for a pair of molecules. For this reason, van der Waals interaction is much more significant for particles than was originally thought.

It is also clear from the short-range expressions for spheres that the energy is directly proportional to the sphere diameter. Although, in general, van der Waals attraction becomes less significant for larger particles, as explained in Section 4.1.1, there are some spectacular exceptions. For instance, it appears that the ability of lizards, such as the gecko, to climb vertical surfaces is a result of “dry adhesion,” which depends primarily on van der Waals forces. The reason is that the gecko’s toes have millions of tiny pads, or *setae*, which give far greater attraction to a surface than when attachment is at a single point, as in the sphere-plate case.

4.2.3 Hamaker constants

Hamaker constants can be calculated in various ways, and, in some cases, they have been derived from direct measurement of attraction forces. Calculations using the original Hamaker method are based on the assumption of complete additivity of intermolecular forces, which is known to be unreliable. An alternative “macroscopic” approach was developed by Lifshitz and co-workers in the 1950s. This makes no assumptions about the molecular nature of the interacting materials and uses only macroscopic properties, in particular, dielectric data. We shall not go into details here, but the Lifshitz result for flat plates gives a result with the same form as the Hamaker expression, Equation (4.5), so that, for spheres at close approach, Hamaker results such as Equations (4.9) and (4.10) should also be of the correct form. It is only the numeric value of the Hamaker constant that differs between the Hamaker (microscopic) and Lifshitz (macroscopic) approaches, and in many cases the results are not greatly different (see Table 4.1).

For nonpolar materials, the major contribution to van der Waals interaction comes from frequencies in the ultraviolet region, and a simple expression is available, based on optical dispersion data. Although this is derived from the definition of Hamaker constant in Equation (4.6), it uses only data obtained from bulk properties of the materials. For the interaction of two

Table 4.1 Calculated Hamaker constants[a]

	A/10^{-20} J			
	In Vacuum		In Water	
Substance	"Exact"	Eq. (4.12)	"Exact"	Eq. (4.16)
Water	3.7	3.9	—	—
Fused quartz	6.5	7.6	0.83	0.61
Calcite	10.1	11.7	2.2	2.1
Sapphire (Al_2O_3)	15.6	19.8	5.3	6.1
Mica	10.0	11.3	2.0	1.9
Polystyrene	6.6	7.8	0.95	0.67
PTFE (Teflon)	3.8	4.4	0.33	0.015
n-Octane	4.5	5.3	0.41	0.11
n-Dodecane	5.0	5.9	0.50	0.21

[a] "Exact" values taken mainly from Israelachvili (1991). Approximate values from Equations (4.12) and (4.16).

materials, 1 and 2, across a vacuum, the Hamaker constant A_{12} is given approximately by the following:

$$A_{12} = \frac{27}{32}\frac{h\nu_1\nu_2}{(\nu_1+\nu_2)}\left(\frac{n_1^2-1}{n_1^2+2}\right)\left(\frac{n_2^2-1}{n_2^2+2}\right) \qquad (4.11)$$

where h is Planck's constant, ν_1 and ν_2 are characteristic *dispersion frequencies* of the materials, and n_1 and n_2 are values of refractive index. The dispersion frequencies are derived from the variation of the refractive index with frequency (typical values are of the order of 3×10^{15} Hz), and the refractive indices are values extrapolated to zero frequency (although values for visible light can be used with very little error).

For the interaction of similar media the Hamaker constant A_{11} is as follows:

$$A_{11} = \frac{27}{64}h\nu_1\left(\frac{n_1^2-1}{n_1^2+2}\right)^2 \qquad (4.12)$$

In most tabulations of Hamaker constants, the values are given for single materials A_{11} (i.e., for the interaction of objects both composed of substance 1). For the interaction of

different materials, the composite Hamaker constant can be calculated approximately from the individual values by the following *geometric mean* assumption:

$$A_{12} \approx \sqrt{A_{11}A_{22}} \tag{4.13}$$

It follows from Equations (4.11) and (4.12) that this approximation would be valid if the dispersion frequencies of the two substances are nearly equal.

For nonpolar materials, there are lower frequency contributions to van der Waals interaction (e.g., as a result of rotation of dipolar molecules). The most important example is water, which has a very high dielectric constant because of the polar nature of water molecules. There is an important "zero frequency" (or "static") contribution to the Hamaker constant, in addition to the "dispersion" component given by Equation (4.12). For water, the zero frequency term is close to $(3/4)k_BT$ or about 3×10^{-21} J, which is less than 10% of the total value of 3.7×10^{-20} J (Table 4.1). However, the zero frequency term can play a much larger part in the interaction of materials through water (see 4.2.4). Another complication is that the zero frequency term is affected by the presence of dissolved salts and is considerably reduced at high ionic strengths.

Some Hamaker constants for various materials of interest are given in Table 4.1. This includes values for materials interacting across a vacuum and in water (see Section 4.2.4). The values given are from "exact" computations based on Lifshitz theory and from approximate expressions, Equations (4.12) and (4.16). Most Hamaker constants are of the order of 10^{-20} J. Higher values apply to fairly dense mineral particles, whereas low-density materials tend to have low Hamaker constants. This is because refractive index values tend to be greater for higher density materials and Hamaker constants depend greatly on refractive index.

Although the values in Table 4.1 may appear to be rather small in terms of energy, they are by no means insignificant. It is reasonable to compare them with a measure of thermal energy, k_BT (where k_B is Boltzmann's constant and T is the absolute temperature). At ordinary temperatures, k_BT has a value of about 4×10^{-21} J, which is of comparable order to Hamaker constants. When the Hamaker constant is 10^{-20} J (about 2.5 k_BT), Figure 4.3 shows that the interaction energy for equal spheres becomes comparable to thermal energy when the separation distance is about 5% of the diameter. At larger separations the interaction would be become insignificant, compared to thermal energy.

4.2.4 Effect of dispersion medium

So far, we have only considered the interaction of objects in a vacuum, but for particles in water we need to extend the treatment to objects separated by another medium (water, in our case). Fortunately, all that is needed is a modified Hamaker constant. In the case of two media, 1 and 2, separated by a third medium, 3, the required constant is $A_{132,}$ given by the following:

$$A_{132} = A_{12} + A_{33} - A_{13} - A_{23} \tag{4.14}$$

where the terms on the right-hand side are Hamaker constants for the interactions between the various media *in vacuo*. Thus, A_{13} represents the interaction between media 1 and 3 across a vacuum, etc.

The form of Equation (4.14) can be explained by the fact that a particle in a suspension effectively displaces an equivalent volume of suspension medium. When particles approach each other from a large distance, new particle-particle and medium-medium interactions are created but two particle-medium interactions are lost (see Figure 4.4). This effect may also be thought of as analogous to the Archimedes principle of buoyancy.

With the "geometric mean" assumption, Equation (4.13), the expression for A_{132} becomes the following:

$$A_{132} = (A_{11}^{1/2} - A_{33}^{1/2})(A_{22}^{1/2} - A_{33}^{1/2}) \tag{4.15}$$

For similar materials, 1, interacting through medium 3, the corresponding expression is as follows:

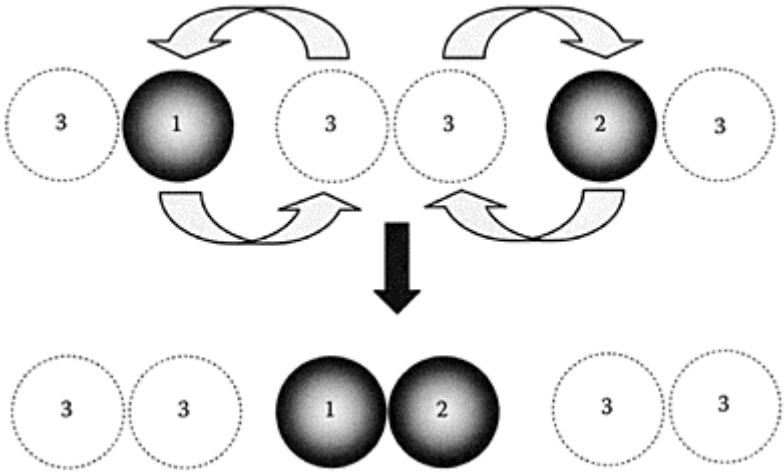

Figure 4.4 Interaction of two particles 1 and 2 in medium 3. In bringing the particles together, equivalent volumes of medium are displaced, as shown. The process involves the loss of one 3–3, one 1–3, and one 2–3 interactions and the gain of two 3–3 and one 1–2 interactions. This reasoning leads to Equation (4.14).

$$A_{131} = (A_{11}^{1/2} - A_{33}^{1/2})^2 \tag{4.16}$$

This shows that interactions between similar media across another medium should always give a positive Hamaker constant (i.e., an attraction). However, for different materials, Equation (4.15) suggests that a negative Hamaker constant, and hence a van der Waals *repulsion*, could arise under certain conditions. This would happen when A_{33} is intermediate between the other two values (e.g., $A_{11}<A_{33}<A_{22}$). This condition may sometimes be found for nonaqueous systems, but, because the Hamaker constant f or water is lower than that for practically all other materials (see Table 4.1), van der Waals interactions in aqueous media should always be attractive.

Equations (4.15) and (4.16) show that the composite Hamaker constant depends on the difference between the square roots of the value for water and the other media. This means that Hamaker constants in water and other media can be significantly smaller than for interactions *in vacuo*, as shown in Table 4.1. Also, for aqueous dispersions, when the value for particles, A_{11}, is not much greater than that for water, A_{33}, then the composite value, A_{131}, can be greatly influenced by the zero frequency term for water.

The geometric mean assumption in Equation (4.13) only applies to dispersion components of Hamaker constants, and not the zero frequency term, so that expressions such as Equation (4.15) would not apply when the intervening medium is water. A simple way around this problem is to calculate the composite Hamaker constant using only values calculated from optical dispersion data and then to add the zero frequency term to the final result. However, because the zero frequency term is reduced as ionic strength increases, there is still some uncertainty associated with low values of the Hamaker constant.

For particles in water, Hamaker constants are mostly in the range of 0.4 to 10×10^{-20} J. Metallic particles have higher values, but these are not common in natural waters. At the low end of the range the major contribution is from the zero frequency component, so the values my not be very reliable.

This method of dealing with interactions across another medium is not necessary in the full Lifshitz (macroscopic) approach, where the effect of an intervening medium is fully included in the theory However, the Hamaker approach, outlined earlier, may give a better appreciation of the underlying physical principles. Also, even in the macroscopic approach, uncertainties arise because of small differences between quantities that are closely similar. The required spectroscopic data may also not be available.

4.2.5 Retardation

Because van der Waals (dispersion) interactions are electromagnetic in nature, they are subject to a relativistic effect known as *retardation*. A fluctuating dipole in one molecule induces a corresponding dipole in another molecule, giving an attraction. If the molecules are quite far apart, a finite time is needed for transmission of the interaction, and, in effect, the fluctuations become out of phase. This leads to a reduced attraction and a different dependence on separation distance. When the molecules are very far apart and the interaction is “fully retarded,” the interaction energy varies as $1/r^7$, rather than $1/r^6$, as

in Equation (4.4). However, at such large separations, intermolecular attraction is negligible, so that retardation is of only minor significance.

Between macroscopic objects retardation *can* give a significant reduction in van der Waals attraction. Although this effect is included in the macroscopic theory, the simpler Hamaker approach can be modified to account for retardation. In many cases, a simple empirical correction factor can be used. For the interaction of spheres, a correction factor can be applied to Equation (4.9):

$$V_A = -\frac{A_{12}}{12h}\frac{d_1 d_2}{(d_1 + d_2)}\frac{1}{(1 + 12h/\lambda)} \qquad (4.17)$$

where λ is a characteristic wavelength of the dispersion interaction, which can generally be assumed to be of the order of 100 nm.

Equation (4.17) gives reasonable agreement with more exact computations and shows that the retardation effect can be significant, even at separations of a few nanometers. At a distance of 10 nm the interaction energy is less than half of the unretarded value.

So, uncritical use of the simple Hamaker expression for spheres, Equation (4.17), can lead to an overestimate of the van der Waals attraction for essentially two reasons:

- For geometric reasons, Equation (4.17) gives results that are too high when *h/d* exceeds about 0.01 (see Figure 4.3).
- The retardation effect gives a significant reduction when *h* is greater than about 1 nm.

These problems arise even when reliable values of the Hamaker constant are available, which is not always the case.

4.3 Electrical double-layer interaction

4.3.1 Basic assumptions

It was seen in Chapter 3 that most particles in water are charged and carry an *electrical double layer*. As two charged particles approach each other in water, the diffuse parts of their double layers begin to overlap and this causes an interaction. For particles with similar charge this gives a repulsion, which is the origin of colloid stability in many cases. A detailed treatment of this subject is beyond the scope of this book; we shall restrict our attention to fairly simple approximations. However, the assumptions involved are usually reasonable for particles in water.

The most important assumption is that the interaction between charged particles depends on the zeta potential, ζ, rather than the "true" surface potential, ψ_0. The electrokinetic, or zeta, potential is believed to be close to the Stern potential ψ_δ (see Figure 3.3). Effectively we assume that the electrokinetic plane of shear, where the potential is ζ (Figure 3.6), coincides with the Stern plane, which represents the closest approach of counterions to the surface.

This assumption has several advantages:

- The zeta potential can be derived experimentally in many cases.

- The zeta potential is considerably lower than the surface potential, which makes some useful approximations more likely to be acceptable.
- Double-layer interactions are predominantly determined by the diffuse layers around particles, so that the zeta potential is much more relevant than the surface potential.

Theoretical treatments of double-layer interaction have dealt with two limiting conditions: *constant potential* and *constant charge*. In the former case, it is assumed that the surface potential ψ_0 remains constant as the surfaces approach each other. When the surface potential is governed by potential-determining ions and the Nernst equation (Equation 3.1), this might seem a reasonable assumption. However, collisions of particles can be very rapid and it may not be possible for equilibrium conditions to be maintained, casting some doubt on the constant potential assumption.

If the particles have a fixed number of charges and hence a constant surface charge density, then the constant charge assumption may seem rea-sonable. However, this assumption leads to some physically unrealistic conditions as surfaces approach very closely.

It is generally accepted that the constant potential and constant charge assumptions represent hypothetic extremes and that some intermediate condition is more appropriate. The approach adopted here is to use an approximate expression, which gives results between the two extremes and is likely to be acceptable for many practical purposes.

We shall first deal with the interaction of charged flat plates, and, via the Deryagin approximation (see Section 4.1.3), derive expressions for the interaction of spheres at close approach. It will be assumed the potentials are low (less than about 50 mV), which is nearly always the case for the zeta potential of particles in water. Another convenient assumption is that solutions contain only symmetric (z–z) electrolytes, such as NaCl and $CaSO_4$. Otherwise, the equations become cumbersome.

4.3.2 Interaction between flat plates and spheres

Consider two parallel flat plates with zeta potentials ζ_1 and ζ_2, both assumed to be quite low, immersed in a symmetric (z–z) electrolyte solution. If the plates are far apart, then the potential distribution in solution adjacent to each plate would be unaffected by the other and would show an exponential decline, according to Equation (3.3). However, as the plates approach more closely, they exert a mutual influence and the potential distribution between the plates adopts a form like that in Figure 4.5, which shows a minimum potential at some distance from the plates.

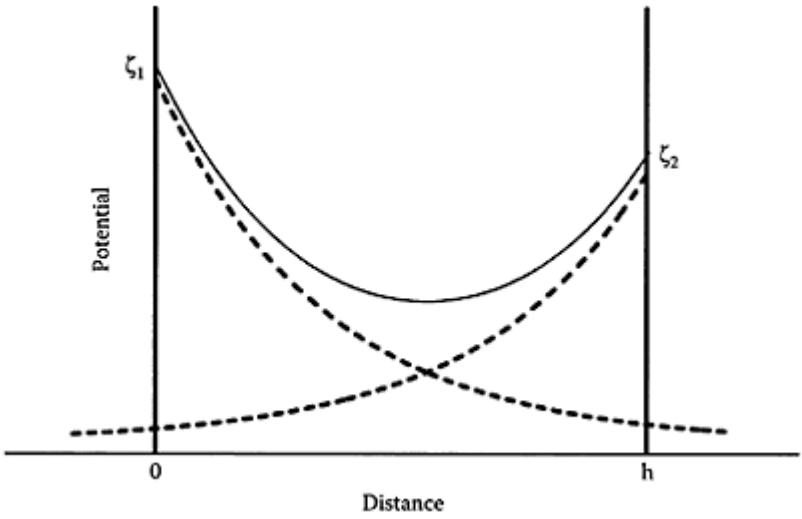

Figure 4.5 Showing the distribution of potential between two flat plates with different (zeta) potentials *(full line),* under the linear superposition approximation (LSA). The dashed lines show the exponential fall of potential from the isolated plates. The potential between the plates is assumed to be just the sum of the isolated plate values.

Provided that the potential between the plates is everywhere low, it can be shown that the force per unit area between the plates is given by the following:

$$P = n_0 k_B T \left[y^2 - \frac{1}{\kappa^2} \left(\frac{dy}{dx} \right)^2 \right] \tag{4.18}$$

where n_0 is the number concentration of cations (or anions) per unit volume and κ is the Debye-Hückel parameter defined in Equation (3.4). The term y is a dimensionless form of the potential, defined as follows:

$$y = \frac{ze\psi}{k_B T} \tag{4.19}$$

where z is the valence of the ions, e is the electron charge, and ψ is the potential at a point between the plates, at a distance x from one of them.

Equation (4.18) and other approximate expressions in this section are reasonably good for y<2, or potentials less than about 50 mV for 1–1 electrolytes, where z=1.

The first term in square brackets in Equation (4.18) is essentially an osmotic pressure, which arises because, when diffuse layers overlap, there is a higher counterion concentration than for isolated plates. The second term is the *Maxwell stress*, which depends on the potential gradient. The pressure is the same anywhere between the plates, so it is convenient to choose the plane where the potential passes though a minimum. Here the potential gradient is zero and there is no Maxwell stress, so the pressure can be calculated just from the osmotic term:

$$P = n_0 k_B T y_{\min}^2 \tag{4.20}$$

It may be assumed that the potential in the region of the minimum is just the sum of the contributions from the isolated plates (see Figure 4.5). This is called the *linear superposition approximation* (LSA) and leads to the following expression for the force between plates:

$$P=2\varepsilon\kappa^2\zeta_1\zeta_2 \exp(-\kappa h) \tag{4.21}$$

where ε is the permittivity of water.

The corresponding expression for the potential energy of interaction is as follows, from Equation (4.1):

$$V_E=2\varepsilon\kappa\zeta_1\zeta_2 \exp(-\kappa h) \tag{4.22}$$

(The term V_E has been used for electrical interaction, to distinguish it from the van der Waals attraction energy, V_A.)

The essential features of Equation (4.22) are that the interaction depends on the product of the zeta potentials of the plates and exponentially on the distance between them. The exponential term contains the Debye-Hückel parameter, κ, which acts as a scaling term. When κ is high (i.e., at high salt concentrations) the interaction is of rather short range, decaying rapidly with distance. At low salt concentrations, where κ is low, the interaction is of longer range. This point is important for colloid stability. For zeta potentials of the same sign, V_E is positive, so the interaction is always repulsive, and for opposite signs the plates should attract each other.

The interaction energy between *two spheres*, with diameters d_1 and d_2, and zeta potentials ζ_1 and ζ_2, separated by a distance h, is as follows, by the Deryagin method (see Section 4.1.3):

$$V_R = 2\pi\varepsilon\zeta_1\zeta_2 \frac{d_1 d_2}{d_1 + d_2} \exp(-\kappa h) \tag{4.23}$$

For a *sphere-plate* system ($d_2=\infty$):

$$V_R=2\pi\varepsilon\zeta_1\zeta_2 d_1 \exp(-\kappa h) \tag{4.24}$$

Although there are far more elaborate treatments of double-layer interaction, the expressions given in this section are reasonable approximations and are adequate for a discussion of colloid stability.

4.4 Combined interaction—DLVO theory

4.4.1 Potential energy diagram

By making the reasonable assumption that van der Waals and electrical double-layer interactions between particles are additive, it is possible to discuss the stability of colloidal particles in a quantitative manner. This approach was taken originally by two research teams working independently—Deryagin and Landau in Moscow, and Verwey and Overbeek in The Netherlands. The outbreak of World War II prevented contact between these groups, which led to some dispute over priority. Nevertheless, their combined efforts are marked by the name *DLVO theory*, which is now widely used.

By taking the previous simple expressions for van der Waals and electrical interactions and restricting attention to equal spheres, of diameter d and zeta potential ζ, the following expression is derived for the *total* interaction energy, V_T:

$$V_T = \pi\varepsilon\zeta^2 d \exp(-\kappa h) - \frac{Ad}{24h} \qquad (4.25)$$

The terms on the right-hand side are from Equations (4.23) and (4.9), for V_E and V_A, respectively, with $d_1=d_2=d$ and $\zeta_1=\zeta_2=\zeta$. Because of the assumptions made, this expression will only apply for fairly low values of ζ, for a relatively close approach $(h << d)$, and at separations where retardation is not significant (h<5 nm). Although these are somewhat restrictive conditions, Equation (4.25) is still useful as a starting point for discussion of DLVO theory.

Figure 4.6 shows the total interaction energy for interacting spheres, diameter 1 μm, as a function of separation distance, h. The electrolyte is assumed to be a 1–1 electrolyte at a concentration of 50 mM; the zeta potential is 25 mV, and the Hamaker constant is 2 k_BT (or about 8.2×10^{-21} J), which is typical of values for particles in water (see Table 4.1). The interaction energy is also expressed in units of k_BT.

Figure 4.6 is the well-known *potential energy diagram*, which is hugely important in understanding colloid stability. The most obvious feature of this diagram is the large potential energy barrier, with a height of about 80 k_BT. Approaching particles would have to have a combined energy exceeding this value to come into contact. Because the barrier height is so much larger than the average thermal energy of particles ($3k_BT/2$), it is extremely unlikely that colliding particles would be able to surmount the barrier. In other words, a suspension under these conditions would be *colloidally stable*.

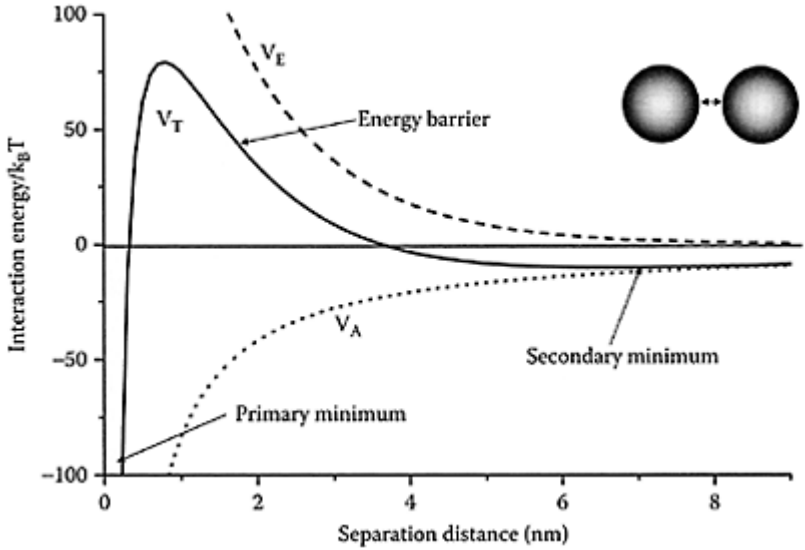

Figure 4.6 Potential energy diagram for the interaction of equal spheres, diameter 1 μm, in a 50-mM solution of 1–1 electrolyte. The zeta potential of the particles is assumed to be 25 mV, and the Hamaker constant is 2 k_BT. The curves show the electrical (V_E), van der Waals (V_A), and total (V_T) interaction energy.

If the potential energy barrier *could* be overcome, then the particles would be held in a deep *primary minimum*. According to Equation (4.25), as *h* tends to zero, the electrical repulsion would reach a finite limit [because exp (0)=1], but the van der Waals attraction would become infinitely large. In practice short-range repulsion forces (not yet discussed) would prevent particles coming into true contact, so the attraction remains finite, although still much larger than the electrical repulsion.

Another important point is that at larger separations there is a shallow *secondary minimum*, which arises because of the different distance dependence of the two types of interaction. Electrical double layer repulsion decays exponentially with distance, whereas van der Waals attraction varies inversely with distance. It follows that, at sufficiently large distance, the attraction term will always be larger than the repulsion, hence the secondary minimum. Whether this minimum is significant relative to thermal energy depends on the particle size and ionic strength, which determines the value of κ and hence the range of repulsion. Usually for particles sizes of around 1 μm or greater and moderate ionic strength, the secondary minimum may have a depth of a few k_BT and

hence be sufficient to hold particles together fairly weakly. This effect may be important in some cases of practical interest.

4.4.2 Effect of ionic strength—critical coagulation concentration

As salt concentration or ionic strength is varied both the zeta potential and the Debye-Hückel parameter κ will be affected. For the present we shall consider the (rather unrealistic) case that the zeta potential remains constant, independent of ionic strength. Then the only effect of ionic strength is on κ, which determines the range of repulsion through the exponential term in Equation (4.25). This effect is known as *double-layer compression* and is shown schematically in Figure 4.7. At low ionic strength the diffuse layer around the particle is extended and particles are prevented from coming into contact. As the salt concentration is increased, the diffuse layer becomes thinner and particles can approach closer before any repulsion is felt. At close approach, van der Waals attraction may be sufficient to outweigh double- layer repulsion.

The potential energy curves in Figure 4.8 are for different concentrations of a 1–1 electrolyte, from 50 to 400 mM. Other conditions are the same as for Figure 4.6. As the salt concentration is increased, the barrier height is reduced and the maximum shifts to smaller distances. This is a direct result of the increase in κ, which causes the repulsion at a given distance to decrease. At a critical concentration (196.6 mM in this case), the maximum occurs at V_T =0. In conventional DLVO theory this is regarded as the salt concentration at which the particles are completely destabilized; it is called the *critical coagulation concentration (ccc)*. It is apparent from Figure 4.8 that, even at this concentration, there is a significant secondary minimum from which parti- cles would have to surmount a barrier of around 25 k_BT to achieve contact in the primary minimum. At about 400 mM salt the barrier has almost disappeared and the particles would have an unhindered path into the primary minimum. For smaller particles, the depth of the secondary minimum at the ccc becomes smaller, and this complication is less of a problem.

The form of the curves in Figure 4.8 shows that the concept of a “critical” concentration is not always clear-cut. Nevertheless, it is worthwhile to establish this concentration in terms of other parameters. All we need to do is find the condition at which $dV_T/dh=0$ and $V_T=0$. It can easily be shown that the maximum then occurs at a separation distance $h^*=1/\kappa$ (i.e., the “thickness” of the diffuse layer). Substituting this value in Equation (4.25) and setting $V_T=0$ gives the value of κ corresponding to the

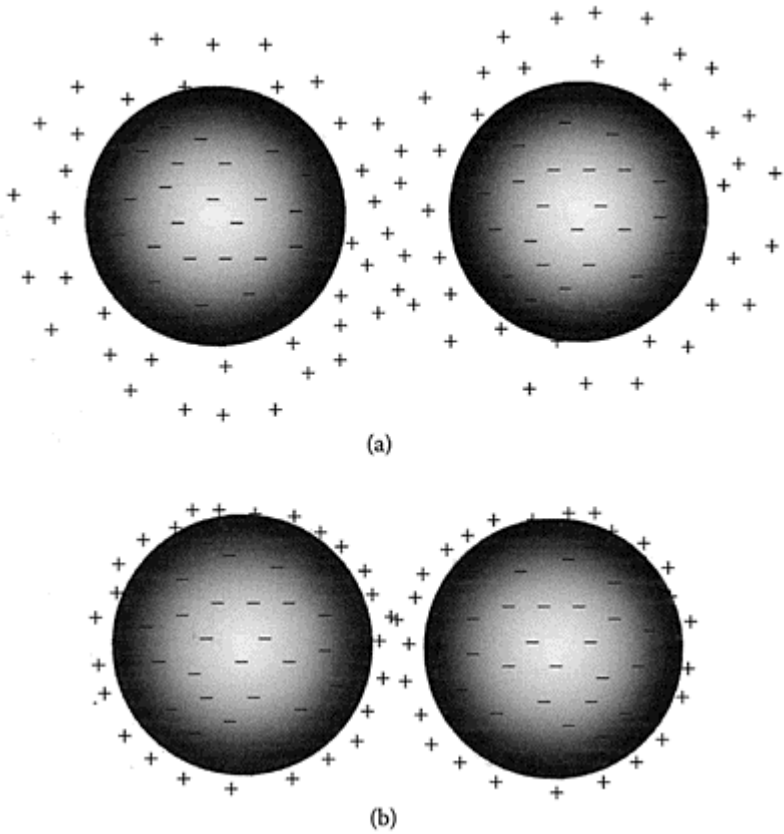

Figure 4.7 Schematic picture of the effect of ionic strength on the range of double-layer repulsion. (a) Low and (b) high salt concentration.

critical coagulation concentration. From the definition of κ in Equation (3.4), the critical concentration is found to be the following:

$$\mathrm{ccc(M)} = 3.41 \times 10^{-35} \frac{\zeta^4}{z^2 A^2} \tag{4.26}$$

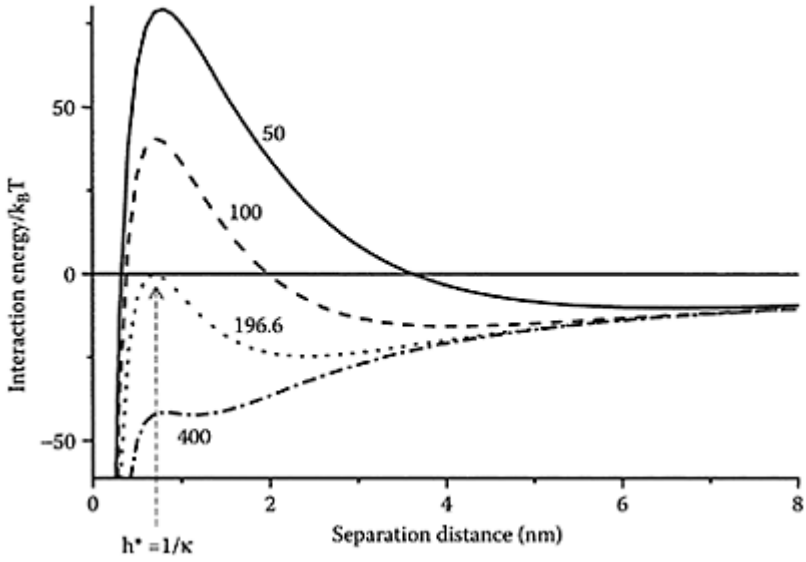

Figure 4.8 The effect of 1–1 electrolyte concentration on the total interaction energy. The electrolyte concentrations (mM) are shown on the curves. All other conditions are as for Figure 4.6.

(The numeric constant is appropriate for aqueous dispersions at 25°C.)

Putting ζ=25 mV, z=1, and $A=2k_BT$ in this expression gives ccc=196.6 mM, which corresponds to the critical concentration shown in Figure 4.8

Equation (4.26) has a number of important features. It predicts that the ccc is proportional to the inverse *square* of the ion charge z. In other words, the ccc for a 2–2 electrolyte should be lower by a factor of 4 (or about 50 mM for the conditions of Figure 4.8). If another expression for double layer repulsion is used, not restricted to low potentials, then an expression for the ccc can be derived that predicts a $1/z^6$ dependence. However, this applies only to the case of very high surface (zeta) potential. This is not realistic—coagulation occurs when zeta potentials are quite low, usually less than about 30 mV, where Equation (4.26) is acceptable.

It has long been known that critical coagulation concentrations depend strongly on ion charge (especially the counterion charge), and the $1/z^6$ dependence is sometimes known as the *Schulze-Hardy rule*. However, this behavior is not often found experimentally. In many cases a dependence nearer to $1/z^3$ is found, so that for a 2–2 electrolyte the ccc would be about 10 times lower than that for a 1–1 electrolyte. It must be remembered that our discussion so far has been entirely concerned with *indifferent* electrolytes, which act entirely through their effect on ionic strength and the Debye-Hückel parameter, κ. No allowance has been made for specific adsorption of counterions (Chapter 3, Section

3.1.4), which is likely with multivalent ions. This point will be discussed further in Section 4.4.3.

It also follows from Equation (4.26) that the ccc depends on the inverse square of the Hamaker constant and on the fourth power of the zeta potential.

It is unrealistic to consider the effects of salt concentration without also allowing for changes in the zeta potential. Generally, as the concentration of indifferent electrolyte is increased, the magnitude of zeta potential diminishes (see Figure 3.6). If the surface charge density of the particle, σ, is known, then the surface (zeta) potential can be calculated from Equation (3.7) and the salt concentration (via the parameter κ). This gives the following expression, where the numeric version is obtained by inserting values of constants appropriate to water at 25°C:

$$\zeta = \frac{\sigma}{\varepsilon\kappa} = \frac{\sigma}{2.28z\sqrt{c}} \qquad (4.27)$$

Here we have assumed a flat surface, which is a reasonable approximation if the diffuse layer is thin $(\kappa d >> 1)$. It is also convenient to assume that the charge density remains constant, independent of salt concentration. This may be acceptable in many cases but is not generally valid. Nevertheless, if σ can be treated as constant, then Equation (4.27) allows us to calculate the zeta potential as a function of salt concentration. Results of such calculations are shown in Figure 4.9 for 1–1, 2–2, and 3–3 electrolytes, assuming a surface charge of 30 mC/m^2 (or about 1 elementary charge per 5 nm^2 of surface). As expected, the zeta potential decreases with increasing salt concentration and with increasing ion charge.

It is also possible, from a rearranged form of Equation (4.26), to calculate, for a given salt concentration, the *critical zeta potential*, ζ*, at which the particles become fully destabilized. This value *increases* as the salt concentration increases because at higher ionic strength the double-layer repulsion is of reduced range and a higher zeta potential is needed to maintain stability. Conversely, at low ionic strength, the diffuse layer is more extensive and a low zeta potential is sufficient to provide the required repulsion. The variation of ζ* with salt concentration is also shown in Figure 4.9.

So, with increasing ionic strength, the value of ζ* increases and the actual zeta potential of the particles falls. It follows that, at a certain salt concentration (the critical coagulation concentration), the two lines intersect. At this concentration, ζ=ζ*. An interesting observation from Figure 4.9 is that the critical zeta potential is the same for all of the salts, independent of z. The value of ζ* is about 26.5 mV in all cases. This value can also be derived from an expression for critical zeta potential, which is obtained simply by combining Equations (4.26) and (4.27):

$$\zeta^* = 4.22x10^5 (\sigma A)^{1/3} \qquad (4.28)$$

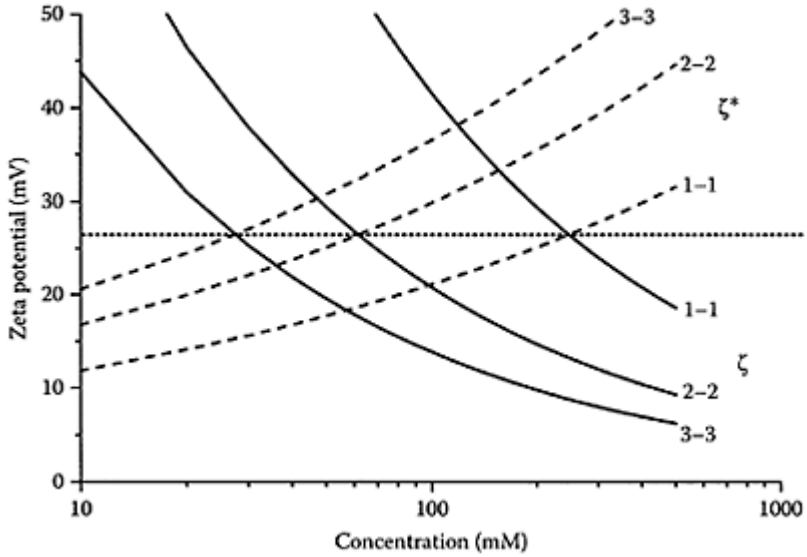

Figure 4.9 Showing the variation of the zeta potential *(full lines)* for particles with a constant surface charge (30 mC/m^2) as a function of concentration of "indifferent" z–z electrolytes, calculated from Equation (4.25). The electrolyte type is shown on the curves. The broken lines show calculated "critical" zeta potentials ζ^*, calculated from Equation (4.24) assuming the same particle size and Hamaker constant as for Figure 4.8. The points where the full and broken lines intersect show the critical salt concentrations and zeta potentials at which complete destabilization occurs.

(Again, the numeric constant applies to water at 25°C.)

There are some experimental data that seem to support the conclusion that ζ^* is independent of z, but there are many others that do not. The assumptions leading to this conclusion, especially a constant surface charge density, may not apply in all cases, so a critical zeta potential independent of electrolyte type should not be taken as a general rule.

Even when allowance is made for changing the zeta potential with ionic strength, the critical coagulation concentration is still dependent on $1/z^2$, just as predicted from Equation (4.26). It is likely that when the ccc shows a greater dependence on counterion valence, then some form of specific adsorption is involved.

4.4.3 Specific counterion adsorption

All of the previous discussion of salt effects has been in terms of indifferent electrolytes, which act in an entirely nonspecific way to reduce the zeta potential of particles and decrease the "thickness" of the diffuse layer. Both of these effects reduce the double-layer repulsion between particles at a given separation distance. There are strong effects of ion charge (especially counterion charge), but the theory predicts that all ions of the same valence should behave in exactly the same way. Thus, salts of calcium and magnesium should have the same critical coagulation concentration for a given colloid. Another important point is that, if salts act only through their effect on ionic strength, then the ccc should not depend on the particle concentration.

There are many important cases where ions adsorb *specifically* on a particle surface. In some cases, this may be the origin of surface charge (see Chapter 3, Section 3.1.4), but, more generally, specific adsorption may significantly modify the double-layer structure (see Figure 3.6). The essential feature of specific adsorption is that it occurs for reasons that are not just electrostatic; there needs to be some other physical or chemical affinity of the ion for the surface. We shall only be concerned with ions that adsorb on *oppositely charged* surfaces (i.e., counterions). In this case, the most obvious evidence of specific effects is that counterions can adsorb beyond the point where complete charge neutralization has occurred, giving *charge reversal* (Figure 3.6).

The most important point regarding colloid stability is that specifically adsorbing counterions can modify the surface charge of particles without appreciable changes of ionic strength. This gives a method of modifying colloid stability, purely by adjusting the zeta potential. By changing the surface charge density, the zeta potential varies according to Equation (4.27). It is then possible to calculate the ccc under given conditions from Equation (4.24). As expected, as the surface charge is reduced, the ccc decreases. Alternatively, we could change the charge density at a fixed ionic strength, until the zeta potential reaches a critical value, giving complete destabilization of the particles. The results in Figure 4.10 show either the critical

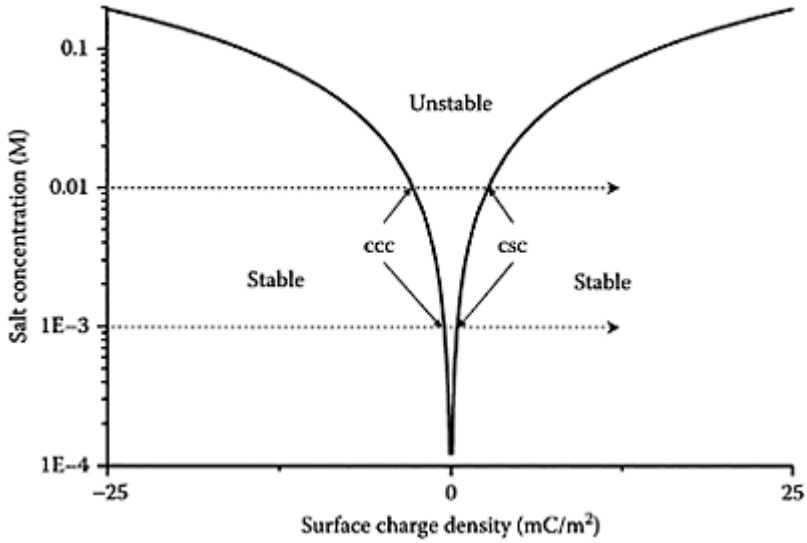

Figure 4.10 Showing the effect of increasing (positive) charge density on colloid stability at different salt concentrations. Coagulation can only occur within a certain range of charge density, which becomes broader as ionic strength is increased. (See text.)

concentration of a 1–1 electrolyte as a function of surface charge density, or, for a given ionic strength, the critical charge densities at which the particles become fully destabilized.

It is clear that, as salt concentration is increased, there is a broader range of charge density that would give full destabilization of the particles. Consider the case where we start with negatively charged particles at a salt concentration of 0.01 M and add small amounts of another salt, containing specifically adsorbing cations. This will reduce the particle charge until, at a value of around −2.6 mC/m^2, the particles are fully destabilized. If we continue to add the specifically adsorbing cations, the particle charge is neutralized and then reversed. When the charge density reaches a value of +2.6 mC/m^2, electrical repulsion between the particles is sufficient to confer stability again. This point is sometimes known as the *critical stabilization condition (csc)*. The term *restabilization* is also used in this context.

At a 10-fold lower ionic strength (10^{-3} M), full destabilization would only occur in a much more restricted range of surface charge density—±0.5 mC/ m^2. This point is highly relevant to the action of certain coagulants (see Chapter 6), especially because typical ionic strengths of natural waters are in the range of 10^{-3} to 10^{-2} M (1–10 mM). At much

higher salt concentrations, the destabilization range becomes very broad and particle aggregation would occur at virtually any charge density

There is no simple way of relating the reduction in charge density to the amount of specifically adsorbing ion added, except for very strong adsorption. In this case, it can often be assumed that the adsorption is *quantitative, i.e.,* that all of the added ions are adsorbed (at least up to the point of charge neutralization). There is then a linear relationship between the amount (dosage) of additive and the reduction in charge density and the *optimum dosage* is strongly related to the original charge density of the particles. It also follows that, in such cases, the optimum dosage is directly proportional to the particle concentration, which is quite unlike the behavior with indifferent electrolytes.

4.4.4 Stability ratio

We have so far only considered complete destabilization when the maximum in the potential energy curve reaches a value V_T=0 (see Figure 4.8). Leaving aside the complication of secondary minima, this point is conventionally regarded as that where the rate of particle aggregation reaches its maximum value (i.e., every collision results in attachment). Conventionally, this is known as *rapid aggregation*, irrespective of the absolute rate. However, even when there is an energy barrier, *some* collisions are effective because a certain fraction of particles will have sufficient energy to overcome the barrier.

The question of aggregation kinetics will be dealt with in the next chapter, but, for the present, we just need to consider *relative* rates of aggregation under the influence of brownian diffusion. Because the aggregation rate reaches a maximum value when the particles are fully destabilized, lower rates are expected when the particles are only partially destabilized. The ratio of the most rapid rate to the rate for a partially destabilized suspension is called the *stability ratio, W*. An equivalent concept is the *collision efficiency*, α, which is the fraction of collisions that result in particle attachment. From these definitions, it follows that

$$W = \frac{1}{\alpha} \tag{4.29}$$

Because these quantities are directly related, it is entirely a matter of convenience whether the stability ratio or the collision efficiency is used in a particular case.

It is possible to relate the stability ratio to the form of the potential energy diagram by treating the problem as one of diffusion in a force field. This leads to the following result for equal spheres, diameter *d:*

$$W = 2\int_0^\infty \frac{\exp\left(V_T/k_BT\right)}{\left(u+2\right)^2}du \tag{4.30}$$

where u is a dimensionless form of the separation distance, $u=2h/d$.

Equation (4.30) involves the term V_T/k_BT because only thermal energy is considered. This approach would not be appropriate if particle collisions were caused by, for instance, fluid motion (see Chapter 5). To evaluate the stability ratio it appears that we

would need to integrate the interaction energy over the entire separation range. However, it turns out that the major contribution to the integral comes from the region around the maximum in the potential energy curve, V_{max}. This allows rough estimates of the stability ratio in terms of V_{max}. For barrier heights of 5, 15, and 25 k_BT, stability ratios are found to be about 40, 105, and 10^9, respectively. For fairly dilute suspensions "rapid" aggregation rates are not high, so that a reduction by a factor of 10^9 would imply essentially indefinite stability.

By considering the effects of indifferent electrolytes on stability ratio, following DLVO theory and with some simplifying assumptions, it is predicted that the stability ratio should depend on salt concentration as shown in Figure 4.11. A plot of log *W* versus log *c* should show two linear regions. Above the critical coagulation concentration, *W*=1 and log *W*=0. At lower concentrations, there should be a linear decline of log *W* with log *c*, and the slope of this line should depend on the zeta potential, particle size, and counterion valence. However, although linear behavior is often found, experimental values of the slope rarely agree with these predictions, and there are several possible explanations, including secondary minimum effects, non-uniform surface charge distribution, and hydrodynamic interaction between particles. Because of these uncertainties, we shall not discuss the variation

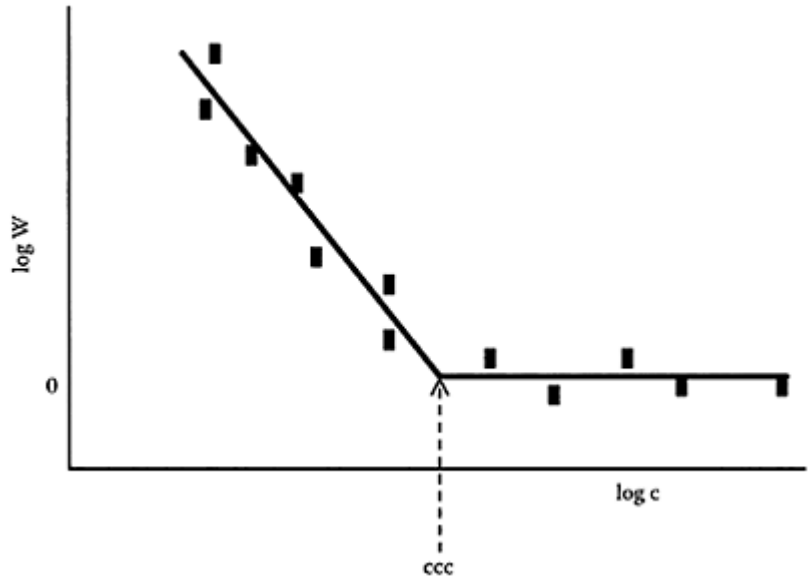

Figure 4.11 Variation of stability ratio, W, with indifferent salt concentration, c. The intersection of lines on a log-log plot gives the critical coagulation concentration.

of W with salt concentration any further. One practical reason for plotting log W against log c is that it provides a convenient way of determining the ccc, as the intersection of the two lines.

4.5 Non-DLVO interactions

Our previous discussion of colloid stability has been entirely in terms of van der Waals attraction and double-layer repulsion. These are the only effects included in the classical DLVO theory All other possible interactions between particles are lumped together under the name "non-DLVO interactions." The most important of these will be considered briefly in the next sections.

4.5.1 Hydration effects

The nature of water close to a particle surface can be very different from ordinary water for various reasons. Because most particles carry a surface charge and hence ionic surface groups, some hydration of these groups would be expected, as for ions in solution. Some particles, especially those of biological origin, have various types of hydrophilic material at their surface, such as proteins and polysaccharides. These can have quite large amounts of "bound water," which plays a part in the interaction of such particles.

The approach of two particles with hydrated surfaces will generally be hindered by an extra repulsive interaction, distinct from electrical double-layer repulsion. This hydration repulsion arises essentially from the need for the surfaces to become dehydrated if true contact between particles is to occur. This involves work and hence an increase in free energy of the system.

Direct measurement of the force between negatively charged surfaces has shown that, at salt concentrations above about 1 mM, an extra short-range force is apparent because of adsorbed hydrated cations. This extra force increases with the degree of hydration of the counterions ($Li^+ \approx Na^+ > K^+ > Cs^+$) and decreases nearly exponentially over the range 1.5–4 nm, with a decay length of the order of 1 nm.

The range of these hydration forces is appreciable compared to the range of double-layer repulsion, and they would be expected to have an effect on colloid stability, especially at high ionic strength. Hydration effects would confer extra stability and may account for some apparently anomalous results.

4.5.2 Hydrophobic attraction

When a surface has no polar or ionic groups or hydrogen-bonding sites, there is no affinity for water and the surface is said to be *hydrophobic*. The nature of water in contact with such a surface will be different from that of ordinary water, which is significantly associated because of hydrogen bonding between the molecules. The cooperative nature of this bonding means that quite large clusters of hydrogen-bonded water molecules can form, although these are of a transient nature, continually forming and breaking down in response to thermal energy fluctuations. Water confined in a gap between two hydrophobic surfaces would be unable to form clusters larger than a certain size. For a

narrow gap, this could be a serious limitation and result in increased free energy of the water in relation to bulk water. In other words there would be an attraction between hydrophobic surfaces. Experiments involving direct surf ace f orce measurements suggest that hydrophobic attraction can be stronger and of longer range than van der Waals attraction. However, there is some evidence that, in these measurements dissolved gas may play a role, with very small bubbles attaching to hydrophobic surfaces and giving significant extra attraction.

For particles dispersed in water, it is possible for their surfaces to have some degree of hydrophobic character. The resulting hydrophobic attraction could be of long range and play a large part in promoting aggregation of particles. There has only been a limited amount of work on the role of hydrophobic effects in colloid stability, and the position is far from clear. There are some indications that removing dissolved air from water can reduce hydrophobic attraction between particles, which provides further evidence that small bubbles are involved. The process known as “oil agglomeration,” where quite large particles, such as coal, are bound together by adhering oil films, is very dependent on the hydrophobicity of the particles. The attachment of bubbles to particles is of crucial importance in the flotation of minerals and is governed by hydrophobic interactions.

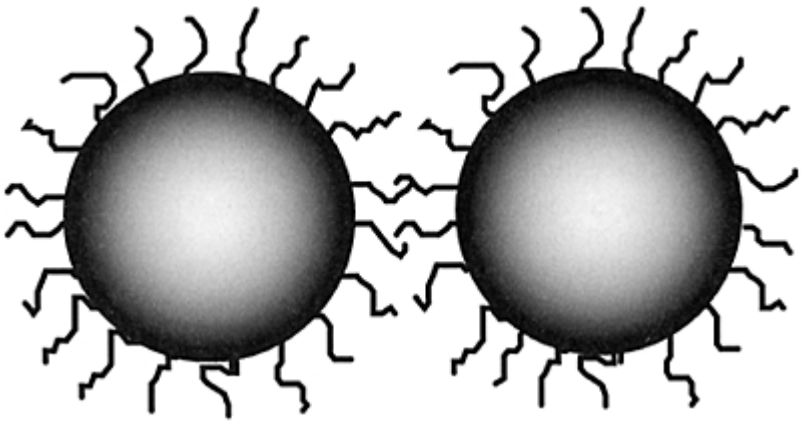

Figure 4.12 Schematic illustration of steric interaction between particles with adsorbed layers of terminally attached polymer chains.

4.5.3 Steric repulsion

Adsorbed layers, especially of polymers, can play a large part in colloid stability. In some cases, small amounts of adsorbed polymer can promote flocculation by a “bridging” mechanism (see later). With larger adsorbed amounts, polymers can give greatly enhanced stability by an effect that is usually known as *steric stabilization*. The most effective stabilizers are polymers that have some affinity for the surface but adsorb in such a way that segments of polymer chains extend some distance into the aqueous

phase. The simplest case is that of terminally adsorbed block copolymers, which have some segments that adsorb strongly on the particles and other, hydrophilic segments that project into the aqueous phase. These polymers form adsorbed layers like those shown schematically in Figure 4.12 and can give greatly enhanced stability. A well-known example for aqueous dispersions is the range of nonionic surfactants, with hydrocarbon segments providing the adsorbing part (by hydrophobic interaction) and hydrophilic "tails" of polyethylene oxide, which can be of various lengths.

The stabilizing action of such materials can be interpreted in f airly simple terms. As particles approach sufficiently close, the adsorbed layers come into contact and any closer approach would involve some interpenetration of the hydrophilic chains. Because these chains are hydrated, overlap of the layers would cause some dehydration and hence an increase in free energy and a repulsion between particles. To a first approximation, the repulsion can be assumed to become infinitely great as soon as the adsorbed layers begin to overlap, but it would be zero at greater separations.

It is reasonable to assume that the effective Hamaker constant for the adsorbed layer is rather low, so that there is little van der Waals attraction between the layers. In that case the most important effect of the adsorbed polymer is to limit the attraction between particles by keeping them a finite distance apart, where van der Waals attraction is lower. Contact between particles may then lead to fairly weak aggregates, which can easily be broken by shear. The most important factor determining the degree of steric stabilization is the thickness of the adsorbed layer relative to the particle size. Because van der Waals attraction energy is proportional to particle size, larger particles will need thicker stabilizing layers to confer the same degree of stability.

Steric stabilization is a widespread phenomenon, and in the older colloid literature materials acting in this way were known as "protective colloids." The classic example is the stabilization of gold sols by gelatin, discovered by Michael Faraday, but the effect has been exploited, unknowingly, since ancient times, as in the preparation of stable dispersions of carbon black (ink) and of other pigments.

In the aquatic environment, most particles have adsorbed layers of natural organic material, such as humic substances, which can have a major effect on their colloidal behavior. Microorganisms produce extracellular polymers that may adsorb and have a great influence on particle interaction in biological systems. These natural polymers and organic materials are often weak acids and are anionic at neutral pH values. It is thought that the rather low negative zeta potentials found for most natural aquatic colloids are a result of the adsorption of this anionic material. The stability of such colloids is often higher than expected on the basis of zeta potential and ionic strength, and it is likely that steric stabilization plays an important part. Humic substances are known to enhance the stability of inorganic colloids and can lead to increased dosages of coagulants in water treatment.

4.5.4 Polymer bridging

Long-chain polymers generally adsorb on particles in the manner indicated in Figure 4.13 and, with large adsorbed amounts, can cause steric stabilization as discussed earlier. With much less adsorbed polymer, an individual chain can become attached to two or more particles, thus "bridging" them together. In this way particles can form aggregates

although they may be charged and repel each other. This effect, originally known as "sensitization," is now widely exploited in many industrial applications.

The essential requirements for polymer bridging are that there should be sufficient unoccupied particle surf ace for attachment of polymer segments from chains attached to other particles and that the polymer bridges should be of such an extent that they span the distance over which interparticle repulsion operates. Generally, the most effective bridging flocculation is found with linear polymers of very high molecular weight (several million), so that extended loops and tails can form, increasing the probability of attachment to other particles. Also, it is found that there is an optimum dosage range for effective flocculation. At lower dosages, there is insufficient polymer to form adequate bridging links between particles. With excess polymer, there is no longer enough bare particle surface for attachment of

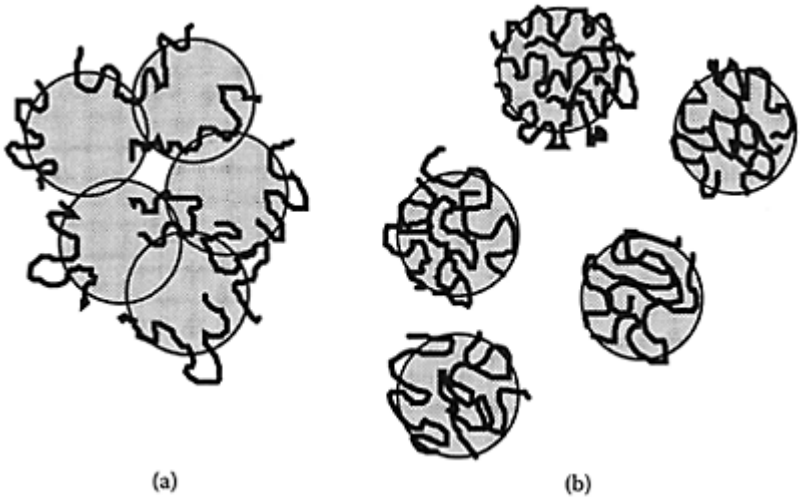

Figure 4.13 Showing (a) bridging flocculation and (b) restabilization by adsorbed polymer chains.

segments and the particles become restabilized, which may involve some steric repulsion. Figure 4.13 shows schematically the flocculation and restabilization of particles by adsorbed polymer. Bridging flocculation can give aggregates (flocs) which are much stronger than those produced just by the addition of salts (i.e., by reducing electrical repulsion). Further discussion of polymeric flocculants will be given in Chapter 6.

Further reading

Churaev, N.V., The DLVO theory in Russian colloid science, *Adv. Colloid Interface Sci*, 83, 19, 1999.

Hunter, R.J., *Introduction to Modern Colloid Science*, Oxford University Press, Oxford, 1993.

Israelachvili, J.N., *Intermolecular and Surface Forces*, 2nd Ed., Academic Press, London, 1991.
Ninham, B.W., On progress in forces since DLVO theory, *Adv. Colloid Interface Sci.*, 83, 1, 1999.
Pashley, R.M. and Karaman, M.E., *Applied Colloid and Surface Chemistry*, Wiley, New York, 2004.

chapter five
Aggregation kinetics

5.1 Collision frequency—Smoluchowski theory

Most discussions of the rate of aggregation start from the classic work of Smoluchowski, from around 1915, which laid the foundations of the subject. It is convenient to think in terms of a dispersion of initially identical particles (*primary* particles), which, after a period of aggregation, contains aggregates of various sizes and different concentrations (e.g., N_i particles of size *i*, N_j particles of size *j*, etc.). Here, N_i and so on refer to the number concentrations of different aggregates, and "size" implies the number of primary particles comprising the aggregate, so that we should think in terms of "i-fold" and "j-fold" aggregates. A fundamental assumption is that aggregation is a second-order rate process, in which the rate of collision is proportional to the product of concentrations of two colliding species. (Three-body collisions are usually ignored in treatments of aggregation; they only become important at very high particle concentrations.) Thus, the number of collisions occurring between *i* and *j* particles in unit time and unit volume, J_{ij}, (the *collision frequency*) is given by the following:

$$J_{ij}=k_{ij}N_iN_j \tag{5.1}$$

where k_{ij} is a second-order rate coefficient, which depends on a number of factors, such as particle size and transport mechanism (see later in this chapter).

In considering the rate of aggregation, it should be remembered that, because of particle interactions, not all collisions may be successful in producing aggregates. The fraction of successful collisions is the *collision efficiency* (see Chapter 4 Section 4.4.4). If there is strong repulsion between particles, then practically no collision results in an aggregate and $\alpha\approx 0$. When there is no significant net repulsion or when there is an attraction between particles, then the collision efficiency will be around unity.

It is usual to assume that the collision rate is independent of colloid interactions and depends only on particle transport. This assumption can often be justified on the basis of the short-range nature of interparticle forces, which operate over a range that is usually much less than the particle size, so that particles are nearly in contact before these forces come into play. However, if there is long-range attraction, then the rate of collision may be enhanced, so that $\alpha>1$.

For the present, we shall assume that every collision is effective in forming an aggregate (i.e., the collision efficiency, α=1), so that the aggregation rate is the same as the collision rate. It is then possible to write the following expression for the rate of change of concentration of *k*-fold aggregates, where *k=i+j:*

$$\frac{dN_k}{dt} = \frac{1}{2}\sum_{\substack{i+j\to k \\ i=1}}^{i=k-1} k_{ij}N_iN_j - N_k\sum_{k=1}^{\infty} k_{ik}N_i \qquad (5.2)$$

The right-hand side of this expression has two terms, representing the “birth” and “death” of aggregates of size *k*. The first gives the rate of formation by collision of any pair of aggregates such that *i+j=k* (e.g., a 5-fold aggregate could be formed by collision of aggregates of sizes 2 and 3 or 1 and 4). The summation procedure in the first term counts each collision twice; hence the factor 1/2. The second term gives the rate of collision of k-fold aggregates with *any other particle* because all such collisions give aggregates larger than *k*.

It is important to point out that Equation (5.2) applies only to *irreversible* aggregation because no allowance for aggregate breakage is made. Breakage of aggregates will be considered later.

The main difficulty of applying Equation (5.2) is finding appropriate values for the collision rate coefficients, k_{ij} and so on. In real systems this is a rather intractable problem and simplifying assumptions have to be made. The coefficients depend primarily on particle size and the mechanisms by which particles collide. Three collision mechanisms are important in practice; these will be discussed in the following section.

5.2 Collision mechanisms

The only significant ways in which particles are brought into contact are as follows:

- Brownian diffusion (*perikinetic* aggregation)
- Fluid motion (*orthokinetic* aggregation)
- Differential sedimentation

These are shown schematically in Figure 5.1 and will be considered in the following sections.

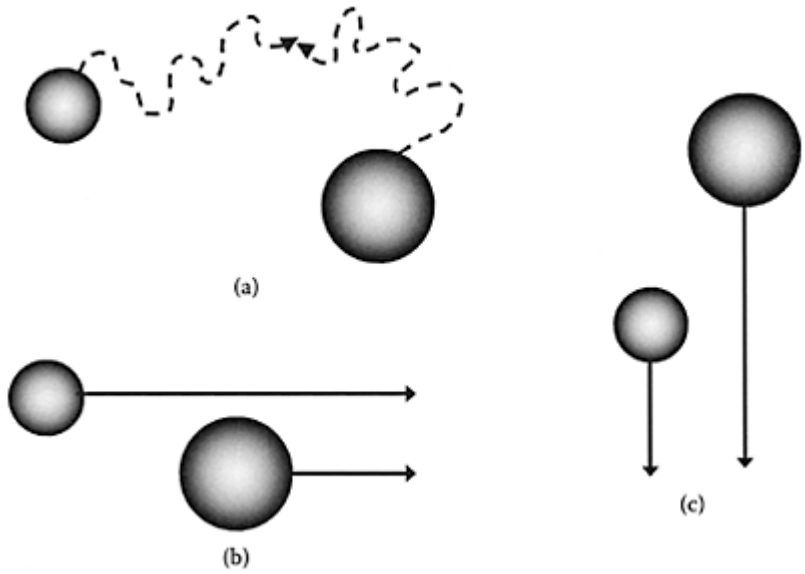

Figure 5.1 Particle transport leading to collisions by (a) brownian diffusion, (b) fluid motion, and (c) differential sedimentation.

5.2.1 Brownian diffusion—perikinetic aggregation

We saw in Chapter 2 (Section 2.3.2) that all particles in water undergo random movement as a result of their thermal energy and that this is known as *brownian motion*. For this reason, collisions between particles will occur from time to time, giving *perikinetic aggregation*. It is not difficult to calculate the collision frequency.

Smoluchowski approached this problem by calculating the rate of diffusion of spherical particles of type *i* to a *fixed* sphere *j*. If each *i* particle is captured by the central sphere on contact, then the *i* particles are effectively removed from the suspension and a concentration gradient is established in the radial direction toward the sphere, *j*. After a very brief interval, steady-state conditions are established, and it can be shown that the number of *i* particles contacting *j* in unit time is as follows:

$$J_i = 4\pi R_{ij} D_i N_i \tag{5.3}$$

where D_i is the diffusion coefficient for *i* particles, given by Equation (2.27), and R_{ij} is the *collision radius*. This is the distance between particle centers at which contact is established. For short-range interactions, the collision radius can be assumed to be just the sum of the particle radii.

Now, in a real suspension, the central particle would not be fixed but would itself be undergoing brownian motion. This is taken care of by using the *mutual diffusion coefficient* for the two particles, which is just the sum of the individual coefficients:

$$D_{ij}=D_i+D_j \tag{5.4}$$

If the concentration of j particles is N_j, then the number of i–j collisions occurring in unit volume per unit time is simply the following:

$$J_{ij}=4\pi R_{ij}D_{ij}N_iN_j \tag{5.5}$$

Comparing this with Equation (5.1) gives the perikinetic collision rate coefficient. If we assume that the collision radius is just the sum of the particle radii and substitute diffusion coefficients from the Stokes-Einstein expression, Equation (2.27), the result is as follows:

$$k_{ij}=\frac{2k_BT}{3\mu}\frac{\left(d_i+d_j\right)^2}{d_id_j} \tag{5.6}$$

This equation has the very important feature that, for particles of nearly equal size, the collision rate coefficient becomes almost independent of particle size. The reason is that the term $(d_i+d_j)^2/d_id_j$ has a value of 4 when $d_i= d_j$ and does not depart much from this value provided that the particle diameters do not differ by more than a factor of about 2. It may seem unreasonable that the brownian collision rate coefficient should not depend on particle size because we know that diffusion becomes less significant for larger particles. However, the collision radius (and hence the chance of collision) increases with particle size, and this effect compensates for the reduced diffusion coefficient. For $d_i\approx d_j$, the rate coefficient becomes the following:

$$k_{ij}=\frac{8k_BT}{3\mu} \tag{5.7}$$

For aqueous dispersions at 25°C, the value of k_{ij} for similar particles is 1.23×10^{-17} m^3/s. For particles of different size, the coefficient is always *greater* than that given by Equation (5.7).

The assumption of a constant value of k_{ij} gives an enormous simplification in the treatment of aggregation kinetics. It is convenient to consider first the very early stages of the aggregation of equal spherical particles. In this case, only collisions between the original single (or *primary*) particles are important, and we can calculate the rate of loss of primary particles just from the second term on the right-hand side of Equation (5.2):

$$\frac{dN_1}{dt}=-k_{11}N_1^2 \tag{5.8}$$

where k_{11} is the rate coefficient for the collision of primary particles, with concentration N_1.

Now, the collision of two single particles leads to the loss of both and the formation of a doublet. So, the net loss in *total* particles (including aggregates) is *one* and the rate of decrease in total number concentration, N_T, is half the rate of decrease in the concentration of primary particles. Thus:

$$\frac{dN_T}{dt} = -\frac{k_{11}}{2} N_1^2 = -k_a N_1^2 \tag{5.9}$$

where k_a is the *aggregation rate coefficient,* which is just half of the collision rate coefficient:

$$k_a = \frac{4k_B T}{3\mu} \tag{5.10}$$

Equations (5.8) and (5.9) apply only to the very early stages of aggregation, where most of the particles are still single. For this reason, they might be thought to be of rather limited use. However, Smoluchowski showed that application of Equation (5.2), with the assumption of constant k_{ij} values given by Equation (5.7), leads to an expression of the same form as Equation (5.9)

$$\frac{dN_T}{dt} = -k_a N_T^2 \tag{5.11}$$

The only difference between this and Equation (5.9) is that N_T, rather than N_1, appears on the right-hand side. This allows integration of the expression to give the total number concentration at time t:

$$N_T = \frac{N_0}{1 + k_a N_0 t} \tag{5.12}$$

Before proceeding further, it is worth remembering that the last two expressions are based on two important assumptions:

- Collisions occur between particles and aggregates not too different in size, so that the collision rate coefficient can be taken as constant.
- Collisions occur between *spherical* particles.

The second assumption is inherent in the Smoluchowski treatment because the problem of diffusion and collision of nonspherical particles is much too difficult to deal with in a simple theory. In reality, although particles may initially be equal spheres, aggregates are unlikely to have spherical

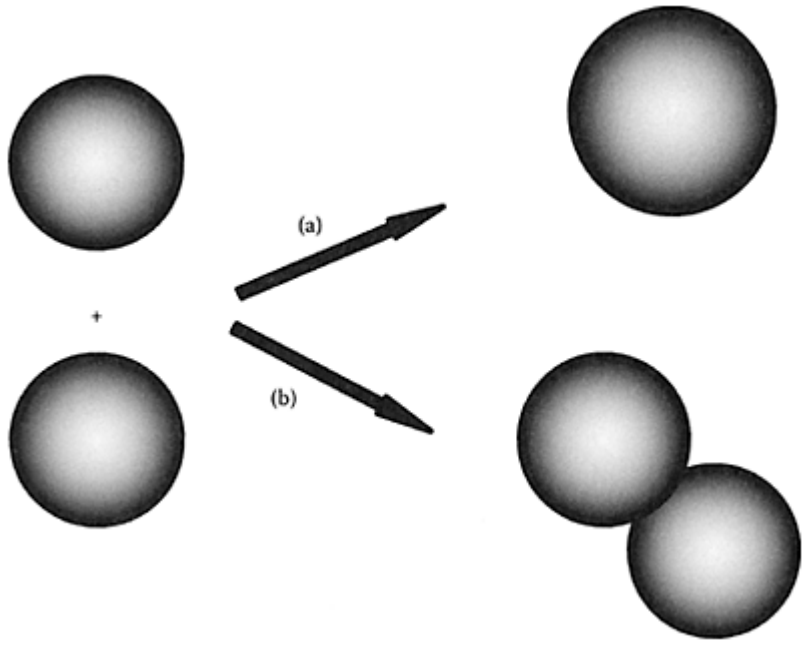

Figure 5.2 Aggregation of two spherical particles to give (a) a larger sphere, by coalescence, and (b) a dumbbell-shaped aggregate.

shape. It is obvious that two hard spheres would collide to give an aggregate with a dumbbell shape (see Figure 5.2), which is clearly nonspherical. The only possibility of a spherical aggregate would be from two colliding liquid droplets (as in an oil-water emulsion) that *coalesce* on contact. We shall return to the question of aggregate shape later (Section 5.3), but for the present it can be assumed that the nonspherical nature of real aggregates does not cause major problems for perikinetic aggregation. Experimental measurements of aggregation rates and aggregate size distributions (see later) are in reasonable agreement with predictions based on the Smoluchowski approach.

It can be seen from Equation (5.12) that the total particle concentration is reduced to half of the original concentration after a time τ, given by the following:

$$\tau = \frac{1}{k_a N_0} \tag{5.13}$$

This characteristic time is sometimes called the *coagulation time* or *half-life* of the aggregation process. This time can also be thought of as the average interval between collisions for a given particle. The fact that τ depends on the initial particle concentration is a consequence of the second-order nature of aggregation kinetics. For first-order

processes, such as radioactive decay, the half-life is completely independent of initial concentration.

Inserting τ in Equation (5.12) gives the following:

$$N_T = \frac{N_0}{\left(1 + t/\tau\right)} \tag{5.14}$$

For aqueous dispersions at 25°C, the value of k_a is 6.13×10^{-18} m^3/s (half of the value quoted previously for k_{ij}). This gives a value of $\tau=1.63\times1017/N_0$. As a numeric example, a suspension initially containing 10^{15} particles per m^3 (or a volume fraction of about 65 ppm for 0.5-μm diameter particles), the aggregation half life would be 163 s. So, in nearly 3 minutes the number concentration would be reduced to 5×10^{14} m^{-3}. To reduce this by a further factor of 2 would require, according to Equation (5.14) a total time of 3τ or about 8 minutes. This example shows that, as aggregation proceeds and the number concentration decreases, longer and longer times are needed to give further aggregation. For this reason, brownian diffusion alone is rarely sufficient to produce large aggregates in a short time (over a period of a few minutes, say). In practice, the rate of aggregation can be greatly enhanced by the application of some kind of fluid shear. This is dealt with in the next section.

Before leaving the subject of perikinetic aggregation, we should mention that the Smoluchowski treatment also allows calculation of the concentration of individual aggregates to be calculated, as well as the total concentration, N_T. The number concentrations of singlets; doublets; and, for the general case, k-fold aggregates can be shown to be the following:

$$N_1 = \frac{N_0}{(1 + t/\tau)^2} \tag{5.15}$$

$$N_2 = \frac{N_0(t/\tau)}{(1 + t/\tau)^3}$$

$$N_k = \frac{N_o(t/\tau)^{k-1}}{(1 + t/\tau)^{k+1}}$$

Results from this expression for aggregates up to 3-fold and for the total number of particles, from Equation (5.15), are shown in Figure 5.3 as a function of dimensionless time, t/τ. Note that for all aggregates the concentration passes through a maximum after a certain time. This is a direct consequence of the “birth” and “death” of aggregates, as given in Equation (5.2). It is also worth pointing out that the concentration of singlets is predicted to exceed that of any individual aggregate at all times. In fact, the concentration of *any* aggregate is always greater than that of any larger aggregate, according to Equation (5.15).

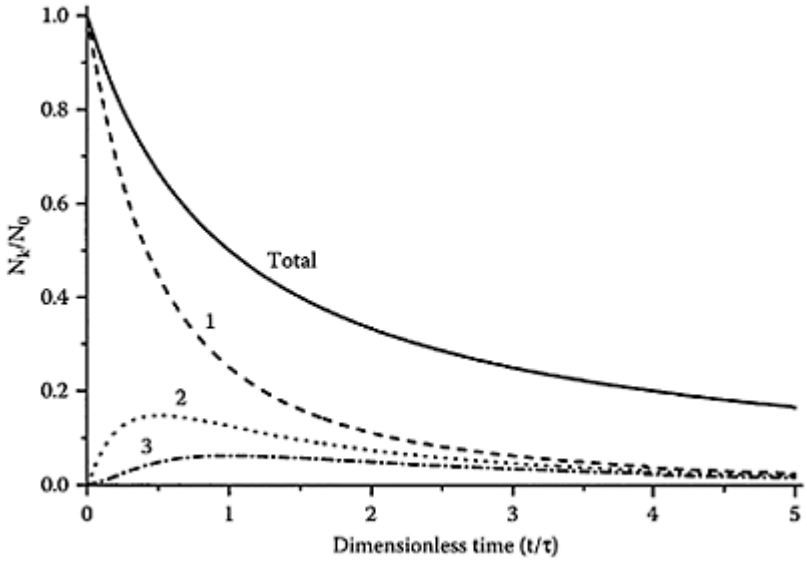

Figure 5.3 Relative concentration of total particle concentration and concentrations of single (primary) particles, doublets, and triplets. Calculated from Equation (5.15).

Although Equation (5.15) is based on various simplifying assumptions, the predicted results in Figure 5.3 are in reasonable agreement with measured aggregate size distributions, when the initial particles are of uniform size.

The aggregate sizes in Equation (5.15) are in terms of the *aggregation number*, *k* (i.e., the number of primary particles in the aggregate). This is proportional to the mass of the aggregate, and so the distribution based on *k* is equivalent to the particle mass distribution discussed in Chapter 2 (although, for aggregates, the mass is usually *not* proportional to the cube of diameter; see Section 5.3). Furthermore, we can define an average aggregation number, $\bar{k}$, which is given by the ratio of the initial number of primary particles, N_0, to the total particle number at some stage in the aggregation process, N_T:

$$\bar{k} = \frac{N_0}{N_T} \tag{5.16}$$

It is then possible to write the aggregation number in a more general, *reduced* form, *x:*

$$x = \frac{k}{\bar{k}} \tag{5.17}$$

The aggregate size distribution is then given as *f(x)*, so that *f(x)dx* is the fraction of aggregates in the (reduced) size range *x*−*x*+*dx*, as in our discussion of particle size distributions in Chapter 2. This approach

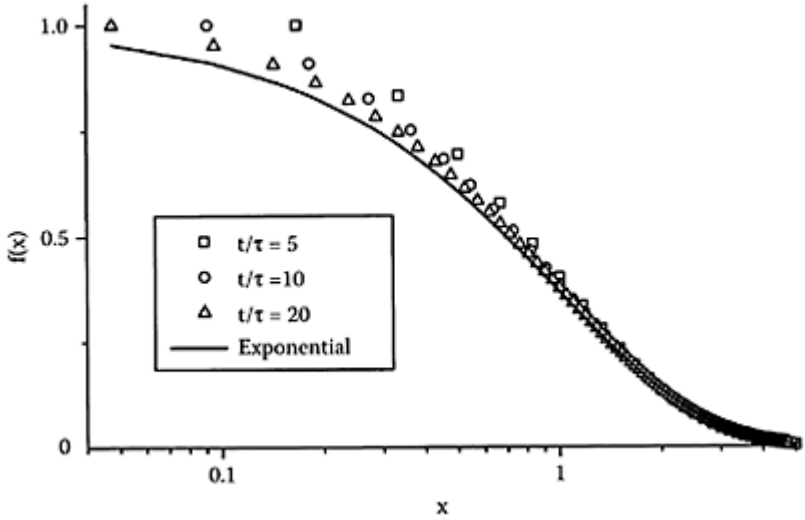

Figure 5.4 Aggregate size distribution plotted as function of the reduced size, x, given by Equation (5.17). The symbols are Smoluchowski results from Equation (5.15) at different aggregation times. The full line is the exponential distribution, Equation (5.18).

assumes a *continuous* size distribution, whereas the Smoluchowski results in Equation (5.15) and Figure 5.3 are in terms of discrete aggregate sizes. However, it is reasonable to equate the fraction of aggregates of size *k*, N_k/N_T, with the frequency function *f(k)*. For any value of the dimensionless aggregation time t/τ, we can calculate the total number of particles N_T from Equation (5.14) and then the average aggregation number from Equation (5.16). It is then possible to calculate the reduced size, *x*, for each aggregate size, *k*, and to plot *f(x)* versus *x*. This has been done in Figure 5.4 for three different aggregation times: 5, 10, and 20τ. Also shown in Figure 5.4 is the exponential distribution:

$$f(x)=\exp(-x) \tag{5.18}$$

The exponential form of aggregate size distribution comes from a *maximum entropy* approach, which finds the most probable distribution of particles among aggregates

without considering details such as collision frequency. It is clear that the predicted distributions from Equation (5.15) are quite close to the exponential form, especially for long times and for aggregate sizes around the average size (x=1) or larger. The discrepancies at small aggregate sizes arise from the fact that Equation (5.15) is a discrete distribution, whereas the exponential form is a continuous distribution. For larger sizes the difference between discrete and continuous distributions becomes less important, and in this region all of the points "collapse" on to the exponential curve. The fact that the Smoluchowski predictions agree quite

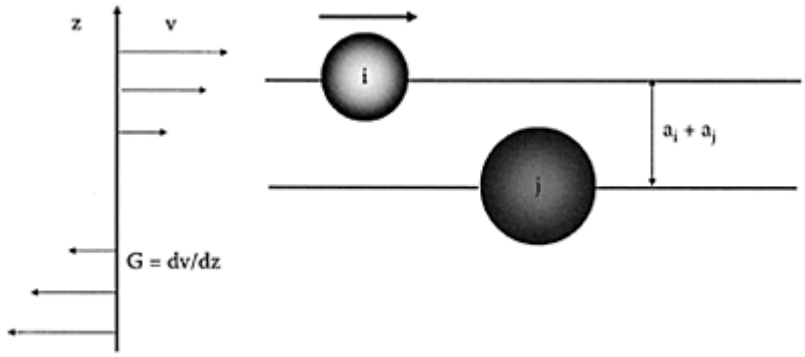

Figure 5.5 Model for orthokinetic aggregation in uniform laminar shear.

well with a distribution derived without considering collision mechanisms is perhaps not surprising because the former are based on the assumption that collision rate coefficients are constant, independent of aggregate size.

The point of the previous discussion is to show that aggregate size distributions, when plotted in a suitable reduced form, can approach a limiting form at long times. These are sometimes known as *self-preserving distributions* because they can emerge in aggregating suspensions, independent of initial conditions (e.g., for nonuniform primary particles). The precise form of the self-preserving distribution depends on a number of factors and may be difficult to predict. Nevertheless, the existence of such distributions can give considerable simplification in theoretical treatments of aggregate size.

5.2.2 Fluid shear—orthokinetic aggregation

We saw in Section 2.1 that brownian (perikinetic) aggregation does not easily lead to the formation of large aggregates because of the reduction in particle concentration and the second-order nature of the process. In practice, aggregation (flocculation) processes are nearly always carried out under conditions where the suspension undergoes some form of *shear*, such as by stirring or flow. Particle transport by fluid motion can have an enormous effect on the rate of particle collision, and the process is known as *orthokinetic aggregation*.

The first theoretical approach to this was also the result of Smoluchowski's work, alongside his pioneering work on perikinetic aggregation. For orthokinetic collisions, he considered the case of spherical particles in *uniform, laminar shear*. Such conditions are never encountered in practice, but the simple case makes a convenient starting point.

Figure 5.5 shows the basic model for the Smoluchowski treatment of orthokinetic collision rates. Two spherical particles, of different size, are located in a uniform shear field. This means that the fluid velocity varies linearly with distance in only one direction, perpendicular to the direction of flow. The rate of change of fluid velocity in the *z*-direction is d*u*/d*z*. This is the *shear rate* and is given the symbol G. The center of one particle, radius a_j, is imagined to be located in a plane where the fluid velocity is zero, and particles above and below this plane move along fluid streamlines with different velocities, depending on their position. A particle of radius a_i will *just* contact the central sphere if its center lies on a streamline at a distance a_i+a_j from the plane where $u=0$. (a_i+a_j is the *collision radius*, as in the treatment of perikinetic aggregation.) All particles at distances less than the collision radius will collide with the central sphere, at rates that depend on their concentration and position (and hence velocity).

It is a straightforward matter to calculate the rate of collision of *i*-particles with the central *j*-particle and hence to derive the frequency of *i*-*j* collisions in a sheared suspension. The result, in terms of particle diameters, is as follows:

$$J_{ij} = \frac{1}{6} N_i N_j G \left(d_i + d_j\right)^3 \tag{5.19}$$

Comparing this with the general rate expression, Equation (5.1), gives the rate coefficient for orthokinetic collisions:

$$k_{ij} = \frac{G}{6} \left(d_i + d_j\right)^3 \tag{5.20}$$

There is an important difference between this and the perikinetic collision rate coefficient, given by Equation (5.6), which becomes almost independent of particle size for roughly equal sizes. By contrast, Equation (5.20) shows a dependence on the *cube* of particle size. This means that, as aggregates form, the decrease in particle number is partly compensated by the increased rate coefficient, so that the aggregation rate does not decline as much as in the perikinetic case.

Because of the strong dependence on particle (aggregate) size, it is *not* possible to assume a constant collision rate coefficient. This makes it much more difficult to predict aggregate size distributions than in the perikinetic case. A numeric approach has to be adopted, but there are still considerable uncertainties in assigning rate coefficients for collisions of aggregates. For this reason, we shall here restrict our attention to the very early stages of aggregation between equal spheres. Assuming that every collision gives an aggregate ($\alpha=1$), we can calculate the rate of decrease of the total particle concentration:

$$\frac{dN_T}{dt} = -\frac{2}{3} N_T^2 G d^3 = -k_a N_T^2 \tag{5.21}$$

This is analogous to the corresponding expression for perikinetic aggregation, Equation (5.11), with the orthokinetic aggregation rate coefficient given by the following:

$$k_a = \frac{2}{3} G d^3 \tag{5.22}$$

Although Equation (5.21) is of much more limited validity than Equation (5.11) and applies only in the very early stages of the aggregation process, it is still useful to explore the consequences for orthokinetic aggregation.

The presence of a d^3 term in Equation (5.21) implies that the volume of particles is a significant factor. The volume fraction of particles in a suspension of equal spheres is just the volume of each particle multiplied by the number per unit volume:

$$\phi = \frac{\pi d^3 N_T}{6} \tag{5.23}$$

Combining this with Equation (5.21) gives the following:

$$\frac{dN_T}{dt} = -\frac{4G\phi N_T}{\pi} \tag{5.24}$$

It might seem reasonable to assume that the volume fraction remains constant during aggregation, so that Equation (5.24) would give a *first-order dependence* of aggregation rate on particle number concentration. Although it is true that the volume of *primary* particles remains constant (assuming no loss by sedimentation), this is not usually the case for aggregates, which have an effective volume greater than that of their constituent particles (see Section 5.3). Nevertheless, it is worth looking at what follows from a "pseudo" first-order rate law.

Assuming that the shear rate, G, and volume fraction, ϕ, remain constant during aggregation, then Equation (5.24) can be integrated to give the following:

$$\frac{N_T}{N_0} = \exp\left(\frac{-4G\phi t}{\pi}\right) \tag{5.25}$$

The exponential term contains the dimensionless group $G\phi t$, which plays an important role in determining the extent of aggregation. In principle, if this group is constant, then the same degree of aggregation will occur whatever the values of the individual terms. For instance, doubling the shear rate and halving the aggregation time should have no effect on the degree of aggregation. In practice, aggregation can be adversely affected at high shear rates (see later), so this conclusion may be misleading.

The dimensionless term *Gt*, sometimes called the *Camp number*, is of great importance in the design of practical flocculation units (see Chapter 6). However, the concentration of particles, ϕ, has equal significance.

All of the previous discussion has been based on the assumption of laminar shear, but this is unrealistic in practice, where aggregation under turbulent conditions is much more common. A way around this problem was introduced by Camp and Stein in 1943. From dimensional analysis they showed that a mean, or effective, shear rate could be derived

from the power input to the suspension (e.g., in a stirred vessel). The mean shear rate can be written in terms of the power input per unit mass, ε, and the *kinematic viscosity, v* (=μ/ρ, where ρ is the density of the suspension):

$$\bar{G} = \sqrt{\frac{\varepsilon}{\nu}} \tag{5.26}$$

The effective shear rate, $\bar{G}$, can be inserted in the previous expressions in place of the laminar shear rate, G. Despite some problems with this approach, it gives results that are in good agreement with more rigorous calculations of collision rates in isotropic turbulence.

5.2.3 Differential sedimentation

Another important collision mechanism arises whenever particles of different size or density are settling from a suspension. Larger and denser particles will sediment faster and can collide with more slowly settling particles as they fall. The appropriate rate can be easily calculated, assuming spherical particles and using Stokes law for their sedimentation rate (see Chapter 2, Section 2.3.3). The resulting collision frequency, for particles of equal density, is as follows:

$$J_{ij} = \left(\frac{\pi g}{72\mu}\right)(\rho_s - \rho_L) N_i N_j (d_i + d_j)^3 (d_i - d_j) \tag{5.27}$$

where *g* is the gravitational acceleration, ρ_s is the density of the particles, and ρ_L the density of the fluid.

It is clear from this equation that the collision rate depends greatly on the size, and also on the *difference* in size, between the colliding particles. The difference in density between the particles and fluid is also important. It turns out that differential sedimentation can become a significant mechanism for large particles and aggregates. Some numeric results will be given in the next section.

5.2.4 Comparison of rates

Three distinct collision mechanisms have now been considered, and it is convenient to compare rates for typical conditions. The most straightforward way of doing this is by comparing the different collision rate coefficients, as defined by Equation (5.1) and the various rate expressions. The appropriate rate coefficients are as follows:

$$\text{Perikinetic: } k_{ij} = \frac{2kT}{3\mu}\frac{(d_i + d_j)^2}{d_i d_j} \tag{5.28}$$

$$\text{Orthokinetic: } k_{ij} = \frac{G}{6}(d_i + d_j)^3$$

$$\text{Differential settling: } k_{ij} = \left(\frac{\pi g}{72\mu}\right)(\rho_s - \rho_L)(d_i + d_j)^3(d_i - d_j)$$

If we consider only collisions between equal spherical particles, diameter *d*, then the differential settling rate is zero. In that case, it is possible to write the ratio of the orthokinetic and perikinetic rate coefficients as follows:

$$\frac{k_{ortho}}{k_{peri}} = \frac{G\mu d^3}{2kT} \tag{5.29}$$

For a shear rate of 10 s^{-1} (corresponding to rather mild agitation), the two coefficients are equal for a particle diameter of about 1 μm. For higher shear rates and, especially, for larger particles, the orthokinetic rate becomes greater.

For unequal particles, the coefficients for all three collision mechanisms are compared in Figure 5.6 for the case of one particle with diameter of 2 μm and a second particle with a diameter varying from 0.01 μm (10 nm) to 20 μm. The shear rate, G, is assumed to be 50 s^{-1}, and the particle density is 2 g/cm^3. All other values needed are appropriate for water at 25°C. There are several noteworthy features of these results:

- The perikinetic rate passes through a minimum when the particle diameters are equal. Around the minimum, the rate coefficient is approximately independent of particle size, as in Equation (5.7). However, when the sizes differ significantly the rate coefficient can be greater than the “constant” value by an order of magnitude or more.
- The differential settling rate is zero for equal particles because they both settle at the same speed and do not collide. However, when the second particle is larger than a few micrometers, this mechanism becomes very important.

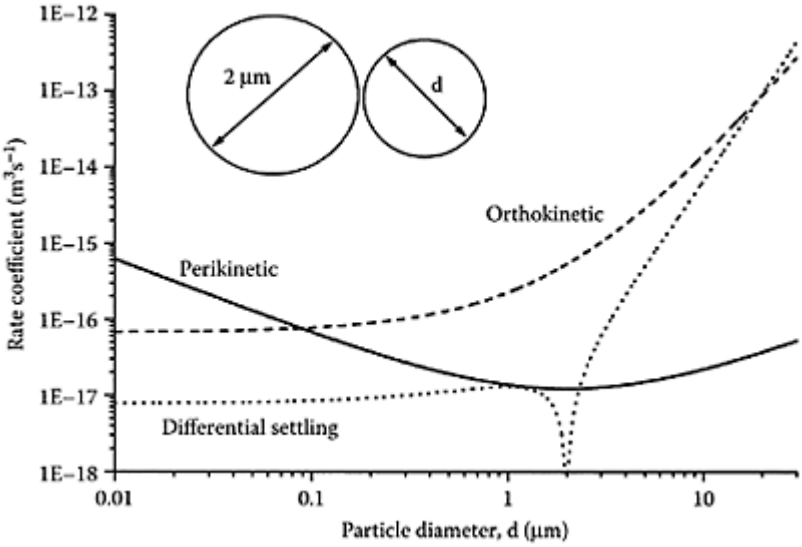

Figure 5.6 Comparison of collision rate coefficients for three different collision mechanisms, as shown. One particle has a diameter of 2 µm and the diameter of the second particle is the variable *d*. The shear rate is 50 s^{-1} and the particle density is 2g/cm^3.

- The orthokinetic rate becomes greater than the perikinetic when the second particle diameter exceeds about 0.1 µm and becomes overwhelmingly larger when *d* is greater than about 2 µm.

Of course, these conclusions would be modified if different values for the fixed particle diameter, density, and shear rate were chosen. However, it can be reasonably assumed that, for particles larger than a few micrometers in agitated suspensions, the perikinetic collision rate will be negligibly small.

5.2.5 Effect of hydrodynamic interactions

All of the previous treatment of aggregation rates has assumed that, for fully destabilized suspensions (α=1), all collisions lead to permanent attachment of particles. In fact, in a viscous fluid, the approach of particles is significantly hindered as the gap between them becomes small. As particles approach close to each other (or to another surface) it becomes increasingly difficult for water to be "squeezed out" of the narrowing gap. This *hydrodynamic* or *viscous* effect tends to slow the aggregation process, even when there is no repulsion between particles.

In the case of collisions induced by brownian motion (perikinetic aggregation), the effect is to reduce the particle diffusion coefficients at close approach. In the absence of any other interactions, this effect would prevent aggregation altogether. However, the universal attractive (van der Waals) forces between particles overcome the viscous resistance to some extent and aggregation can occur, albeit at a reduced rate. For typical values of Hamaker

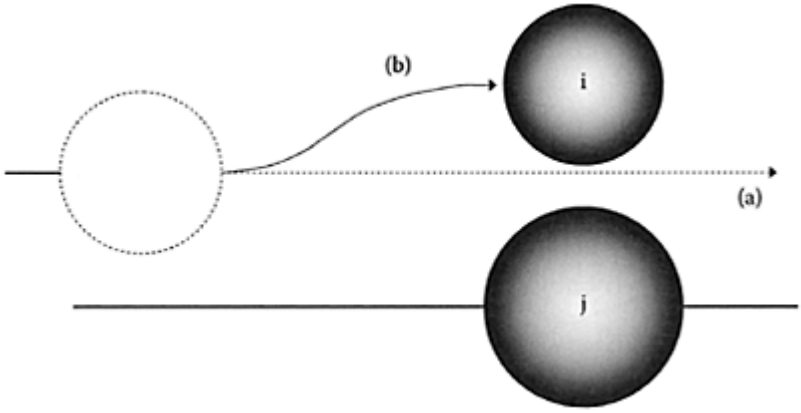

Figure 5.7 Relative motion of two particles in laminar shear. (a) Rectilinear path, as assumed in Smoluchowski theory; (b) Actual, curved path followed in a viscous fluid.

constant (see Chapter 4, Section 4.2.3), the effect is to reduce the Smoluchowski rate, given by Equation (5.11), by a factor of about two.

In the case of orthokinetic collisions, the hydrodynamic effect can be more significant. The problem here is that the model in Figure 5.5 assumes that a particle moves along a straight fluid streamline until collision with another particle occurs. In reality, flow occurs *around* particles, so streamlines must deviate from rectilinear paths (see Figure 5.7). For spherical particles, this effect can be treated numerically, and the results can be expressed as a *limiting collision efficiency*, α_0. This is the fraction of collisions given by the Smoluchowski approach, Equation (5.19), which result in particle attachment, in the absence of any other repulsion between particles (i.e., for a fully destabilized suspension). If there *is* repulsion between particles, such as double-layer repulsion, then the actual collision efficiency will be less than α_0.

Because van der Waals attraction plays an important part in overcoming hydrodynamic resistance, α_0 depends on the value of Hamaker constant. Assuming complete destabilization and a Hamaker constant of 10^{-20} J (about 2.5 kT), the results in Figure 5.8 are found. This figure shows the limiting collision efficiency, α_0, as a function of shear rate f or different values of particle diameter. The fluid is assumed to be water at

25°C. Over a range of G values from 1 to 1000 s^{-1}, the value of α_0 decreases significantly for a given particle size. Also, even at low shear rates, the limiting collision efficiency can be quite small. For instance, aggregation of 10-μm particles at shear rates around 50 s^{-1} would only occur at about 10% of the rate predicted by Equation (5.21).

These results are for particles of equal size. For spheres of different size, collision efficiencies can be much lower still. In some cases, the trajectory of a small particle around a larger one is such that a large separation distance is maintained, such that collision would be very unlikely. For instance, for particles with diameters 20 μm and 2 μm, the distance of closest approach is about 1.4 μm. This is far greater than the range of significant van der Waals attraction, and it is difficult to see how any aggregation could occur.

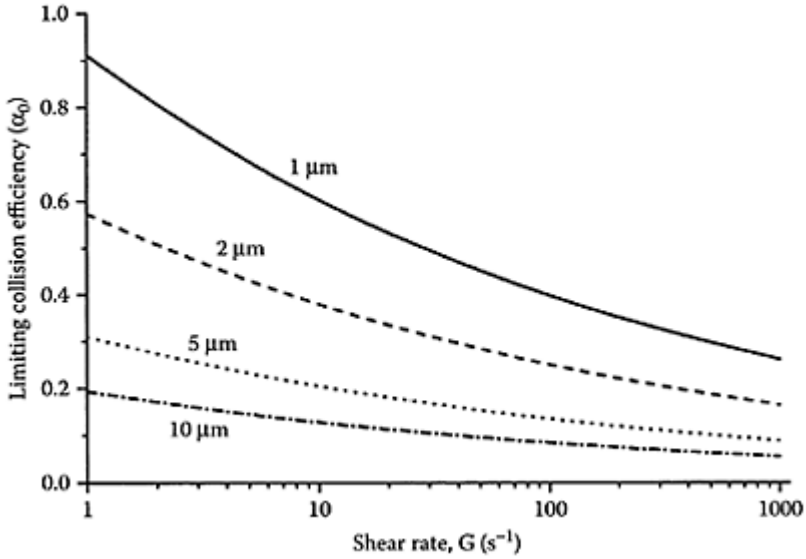

Figure 5.8 Limiting collision efficiency (see text) as function of shear rate. The collisions are for equal spheres, with diameters shown on the curves.

At first sight it would appear that hydrodynamic interaction would severely limit the formation of large aggregates in shear flow. However, there is no doubt that, in practice, large aggregates *can* be formed and that the rate of aggregation is enhanced by increasing the shear rate. Some possible reasons why hydrodynamic effects may not be as important as theoretically predicted are as follows:

- Computations are for hard impermeable spheres in laminar shear flow.
- Effects with irregular particles may be very different.
- The effect of turbulence is not considered.

• Real aggregates often have open structures and may have significant permeability, which would greatly reduce hydrodynamic resistance.

The last point is especially important, and the form of aggregates will be considered in the next section.

5.3 Form of aggregates

When solid particles aggregate, no coalescence can occur and the resulting clusters may adopt many different forms. In the simplest case of equal spheres, there is no doubt about the shape of a doublet, which must be in the form of a dumbbell (Figure 5.2). However, a third particle can attach in several different ways and with higher aggregates the number of possible structures rapidly increases, as indicated in Figure 5.9. In real aggregation processes, aggregates containing hundreds or thousands of primary particles

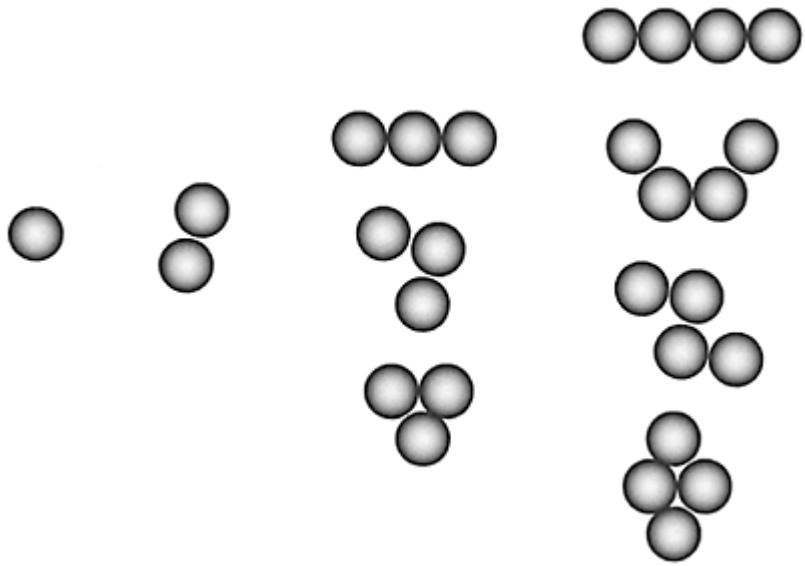

Figure 5.9 Possible forms for aggregates of equal, hard spheres, up to 4-fold.

can form, and it would be impossible to provide a detailed description of their structure. Some convenient method is needed that enables aggregate structure to be characterized in general terms but still conveys useful information. A great deal of progress was made in this area during the 1980s, largely as a result of computer simulation of aggregate formation and the study of model aggregates.

5.3.1 Fractals

Aggregates are now recognized as *self-similar, fractal* objects. Self-similarity means that an aggregate appears to have a similar structure, independent of the scale of observation. A schematic picture of self-similarity in two dimensions is given in Figure 5.10, Here the basic +−shape is maintained over four levels of structure, from individual crosses (level 1), through to the +−shaped arrangement of 125 crosses (level 4). This process could be continued *ad infinitum.* Starting from a single cross (level 1), we can count the number of crosses at each of the stages 1–4 in Figure 5.10. The sequence would be: 1, 5, 25, 125, etc. We can also measure the "size" of each of the individual structures. A convenient measure of the size is just the number of crosses across the middle of each arrangement, giving 1, 3, 9, 27, etc. We now have measures of a linear dimension, L, of each of the structures in Figure 5.10 and the number of crosses in each, N. At each stage, N increases by a factor of 5 and L increases by a factor of 3. It follows that N is related to L by a power law with an exponent given by (log 5)/(log 3)=1.465. Thus:

$$N=L^{1.465} \tag{5.30}$$

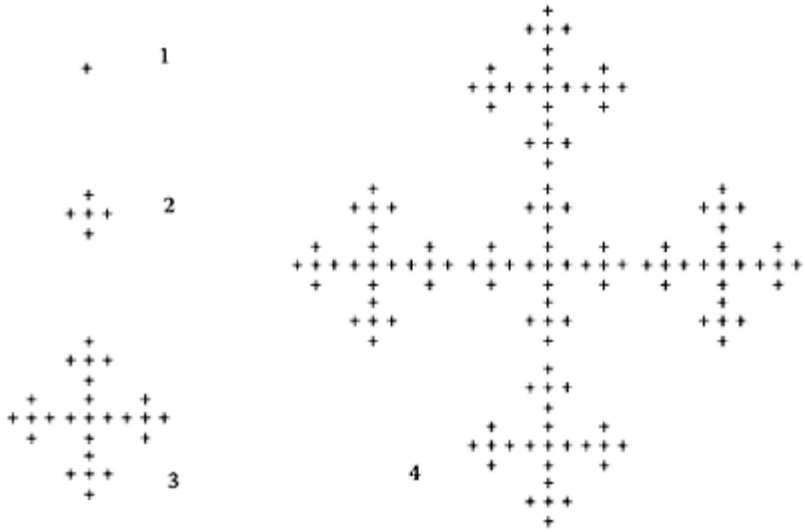

Figure 5.10 Formation of two-dimensional fractal aggregates (see text).

If a two-dimensional plane is filled uniformly with objects, the number of objects should increase as the area of the plane and hence as the *square* of a linear measure. The exponent in Equation (5.30) has a noninteger value, indicating a fractional dimension, so that the objects depicted in Figure 5.10 have *fracfal* character, with a *fractal dimension* of 1.465. This means that, as fractal objects grow larger, they become progressively more open in structure, an effect that is clearly apparent in Figure 5.10.

Fractal, self-similar objects are common in nature. Structures as varied as cauliflowers, trees, and lungs all show branching structures with distinct fractal character. A well-known example is the coastline of an island, such as Britain, where the total length depends on the length of the measuring stick used, and hence on the size of the bays, inlets, and so on, included.

Aggregates of particles have fractal character, such that the mass of an aggregate (or the number of primary particles within it) scales as the size raised to some power, d_F (the *mass fractal dimension*):

$$M \propto L^{d_F} \tag{5.31}$$

It follows that a log-log plot of M versus L should give a straight line, with slope d_p

Aggregates effectively have fractional dimensions; hence the term “fractal,” just as in the two-dimensional case earlier. In three-dimensional space, aggregate fractal dimensions could, in principle, take values anywhere between 1 and 3. A value of d_F=1 would imply a linear aggregate, with a mass proportional to the length. For an aggregate of uniform, nonfractal character, the mass should vary as the cube of the size, giving d_F=3, just as for solid objects.

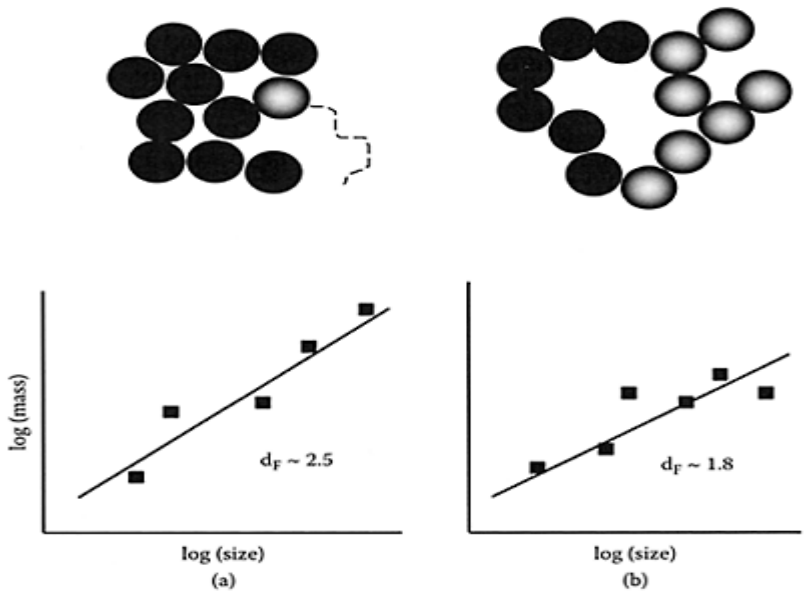

Figure 5.11 Formation of fractal aggregates by (a) particle-cluster and (b) cluster-cluster aggregation.

Early models of aggregate structure were based on the addition of single particles to growing clusters by diffusion (Figure 5.11a). This tends to give quite compact aggregates

with a fractal dimension of around 2.5. Later simulations allowed for cluster-cluster collisions, which is a more realistic model in many practical cases. This gives much more open aggregate structures (Figure 5.11b), with d_F of about 1.8. The reason for the different structures should be clear from Figure 5.11. In the particle-cluster case, the approaching particle is able to penetrate some way into the cluster before making contact. When two clusters approach each other it is likely that contact will occur before the clusters have interpenetrated significantly, giving a much more open structure.

These models were based on the assumption of no repulsion between particles and that collisions occur only by brownian motion (perikinetic aggregation), so that the process is sometimes known as *diffusion-limited aggregation (DLA)*. If there is significant interparticle repulsion, so that the collision efficiency is reduced, then aggregation is said to be *reaction limited*, and more compact aggregates (higher d_F) are formed under these conditions. Again, there is a fairly simple explanation for this effect. Because of the low collision efficiency, particles and clusters have to collide several times before attachment, giving more opportunity to explore different configurations and to achieve some degree of interpenetration.

Aggregates produced under conditions of fluid shear tend to have rather higher fractal dimensions than those formed by diffusion. In stirred suspensions, significant restructuring of aggregates can take place, perhaps by breakage and reformation, which gives a further increase in d_p usually to values of around 2.3.

An earlier concept of aggregate structure was of a "hierarchy" of subunits, ranging from single particles through small aggregates, clusters of aggregates, clusters of clusters, and so on. This idea is well represented, in two dimensions, by structures like those in Figure 5.10, where the hierarchy extends to four levels of aggregation. It is doubtful whether real fractal aggregates follow such a neat pattern, and it is better to think in terms of a continuum of levels rather than discrete values.

Experimental measurement of fractal dimensions of real aggregates is not straightforward, and the methods available are almost always indirect. A light-scattering method is applicable when the primary particles are small enough, and with sufficiently low refractive index, to permit use of the Rayleigh-Gans-Debye (RGD) approximation (see Chapter 2, Section 2.4.6). When the scattered light intensity from a suspension of aggregates is plotted against the scattering vector, q, defined in Equation (2.45), a characteristic pattern is found, as shown in Figure 5.12. At sufficiently low values of q (e.g., at low scattering angles), $1/q$ is much greater than the size of the aggregates and the scattered light intensity depends on the weight-average aggregate size. So, for a given degree of aggregation, the intensity is constant, independent of q. For very large values of q the length scale probed is smaller than the size of the primary particles and the scattered light intensity is just the same for aggregates as that from the isolated primary particles. Again, this leads to a constant scattering intensity, independent of q. For intermediate values of q, such that the length $1/q$ lies between the size of the primary particles and the aggregate size, it can be shown that the scattered light intensity $I(q)$ varies with q according to the following:

$$I(q) \propto q^{-d_F} \tag{5.32}$$

So, on a log-log plot there should be a linear region, with a slope of $-1/d_F$. This method has been widely used, even in cases where the RGD approximation would not be expected to apply. In such cases, a linear region of a log *I(q)* versus log *q* plot is often found, but it is doubtful whether the slope gives a true fractal dimension. A noncommittal term such as "scattering exponent" is sometimes used, and the slope may still give useful empirical information on aggregate structure.

5.3.2 Collision rate of fractal aggregates

The Smoluchowski treatment of aggregation kinetics is based on the assumption that the colliding particles are spheres. Even for spherical primary

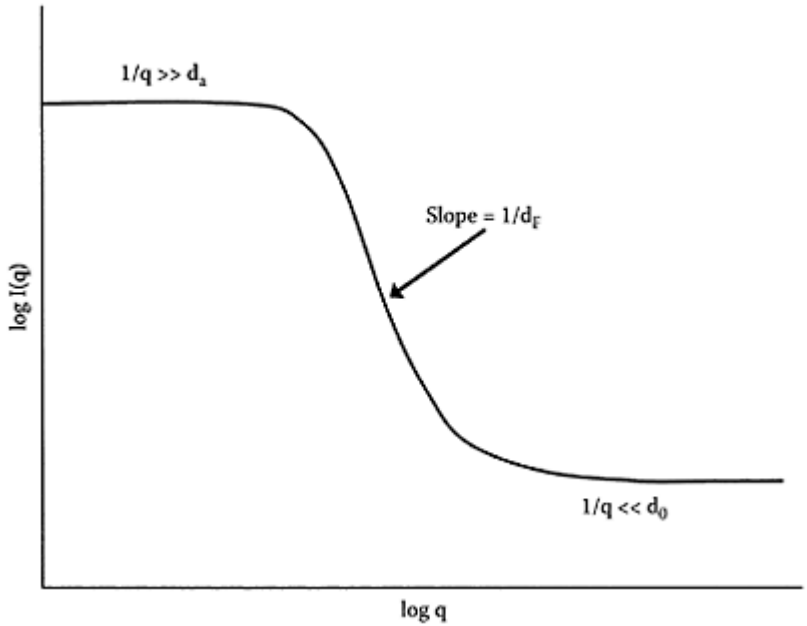

Figure 5.12 Log-log plot of scattering intensity as a function of scattering vector *q*.

particles, aggregation quickly leads to shapes like those in Figure 5.9 and their collision rates cannot be calculated exactly. Only in the case of coalescing liquid droplets could the assumption of spherical particles be justified.

For perikinetic aggregation, the growth of aggregates gives an increasing collision radius and a reduced diffusion coefficient and these effects tend to cancel out, giving a collision rate coefficient that is not greatly dependent on aggregate size (see Section 5.2.1). For fractal aggregates, the hydrodynamic radius (which determines the drag and hence the diffusion coefficient) is likely to be somewhat less than the outer "capture

radius," corresponding to the physical extent of the aggregate. This means that brownian collisions will occur more rapidly than predicted from the rate coefficient in Equation (5.6). However, for aggregates greater than a few micrometers in size, perikinetic aggregation is negligible and collisions induced by shear become far more important.

In the orthokinetic case, it is the effective capture radius of a fractal aggregate that is of paramount importance, and this is greatly dependent on the fractal dimension. Instead of Equation (5.20), the collision rate coefficient for orthokinetic collisions between *i* and *j* particles can be written as follows:

$$k_{ij} = \frac{G\, d_0^3}{6}\left(i^{1/d_F} + j^{1/d_F}\right)^3 \tag{5.33}$$

where d_0 is the diameter of the primary particles. It has been assumed that the diameter of an *i*-fold aggregate is given by the following:

$$d_i = d_0\, i^{1/d_F} \tag{5.34}$$

by analogy with the definition of fractal dimension in Equation (5.31).

For the "coalesced sphere" assumption, d_F=3, and the increase in aggregate size is relatively slow (a 10-fold increase in capture radius for 1000-fold aggregates). For lower values of d_p the aggregate size increases more rapidly, which can give a dramatic increase in aggregation rate. An obvious consequence of the fractal nature of aggregates is that the effective aggregate volume will not be conserved, as assumed in the derivation of Equation (5.25). There will be a substantial increase in effective floc volume for typical values of d_F and this is the reason for the increased collision frequency

Another important consequence of the fractal nature of aggregates is that hydrodynamic interaction is much less significant than for solid particles and the very large effects expected for unequal-size particles are not found for aggregates. It should be clear from the simple pictures in Figure 5.11 that particle-cluster and cluster-cluster collisions would be much less hindered by hydrodynamic effects than similar encounters involving solid particles equivalent in size to the clusters.

5.3.3 Density of fractal aggregates

One of the most important practical implications of fractal structure is that the density of aggregates decreases with increasing aggregate size. This has important consequences for solid-liquid separation processes (Chapter 7). Actually, such behavior had been empirically observed well before the concept of fractal aggregates was introduced.

The effective, buoyant density of an aggregate in a liquid, ρ_E, is simply the following:

$$\rho_E = \rho_A - \rho_L = \phi_S\left(\rho_S - \rho_L\right) \tag{5.35}$$

where ρ_A, ρ_L, and ρ_S are the densities of the aggregate, the liquid and the solid particles, respectively, and ϕ_S is the volume fraction of solid *within the aggregate* (not to be confused with the volume fraction of particles in a suspension).

Experimental values of aggregate density (usually obtained by sedimentation methods) give results like those in Figure 5.13. On a log-log plot of effective aggregate density against some measure of aggregate size (usually diameter) the results often show a linear decline, with a characteristic slope. This implies a relationship of the form:

$$\rho_E = Bd^{-y} \tag{5.36}$$

where B and y are experimental constants.

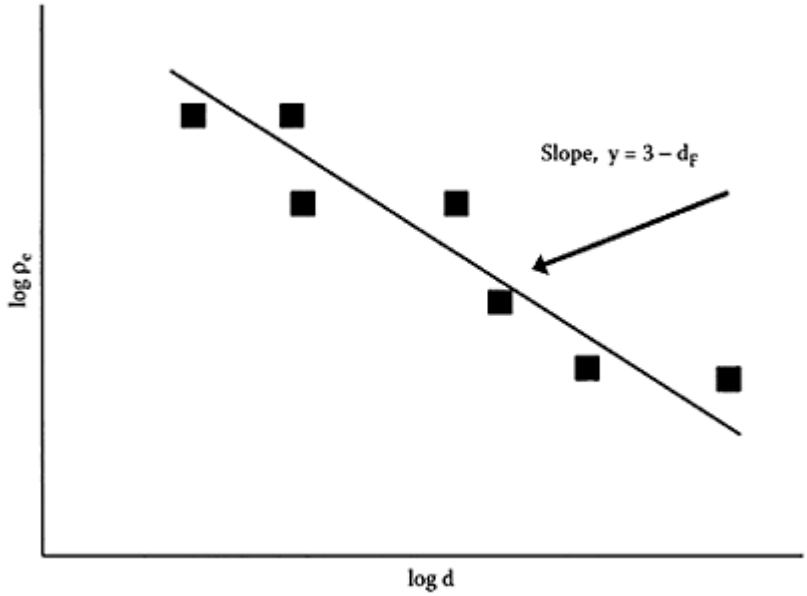

Figure 5.13 Log-log plot of effective aggregate density against aggregate size.

It can easily be shown that ϕ_S is proportional to $d^{(3-d_F)}$ and, because the effective density is directly proportional to ϕ_S, it follows that

$$d_F = 3 - y \tag{5.37}$$

When $d_F=3$ (i.e., for nonfractal objects), $y=0$, so the density does not vary with size, as expected. The lower the fractal dimension, the greater the value of y and hence the more rapid decrease of density with aggregate size. Experimental measurements of aggregate density usually give y values in the range of 1–1.4, corresponding to fractal dimensions of 2–1.6.

5.4 Strength and breakage of aggregates

All of our previous discussion on aggregation kinetics was based on the assumption of irreversible aggregation, so that no breakage of aggregates was allowed. In practice, this is unrealistic because most aggregation processes are carried out in agitated suspensions, nearly always under turbulent conditions. It is inevitable that some breakage of aggregates will occur in these cases.

Empirically, it is found that aggregates grow only to a certain limiting size, which depends on their strength and on the effective shear rate, or energy dissipation rate, ε:

$$d_{max}=C\varepsilon^{-n} \tag{5.38}$$

where C and *n* are constants, depending on the system.

Before discussing aggregate breakage, it should be noted that this is not the only possible reason for the limiting size. We saw in 5.2.5 that hydrodynamic resistance can seriously limit orthokinetic collisions between large particles. Whereas this effect may not be so significant for fractal aggregates, it is still likely that, as aggregates grow larger, further growth may be restricted by a reduced collision efficiency, especially at high shear rates. This might give a limiting aggregate size, even without breakage. However, it is more usual to think of the limiting size as a result of a dynamic balance between aggregate growth and breakage, leading to a steady-state aggregate size distribution.

There are two recognized modes of aggregate breakage:

- *Erosion* of small particles from the surface of an aggregate
- *Splitting* of an aggregate into roughly equal fragments

These are illustrated schematically in Figure 5.14. The mode of breakage depends on the size of the aggregates relative to the turbulence microscale. Turbulence is a complex phenomenon, but it is characterized by eddies of various sizes. For a stirred vessel, the largest eddies are comparable in size to the vessel or impeller. The energy in these large-scale eddies cascades through eddies of decreasing size. The *Kolmogorov microscale* separates the *inertial range*, where energy is transferred with very little dissipation, from the *viscous subrange*, where energy is dissipated as heat. The Kolmogorov

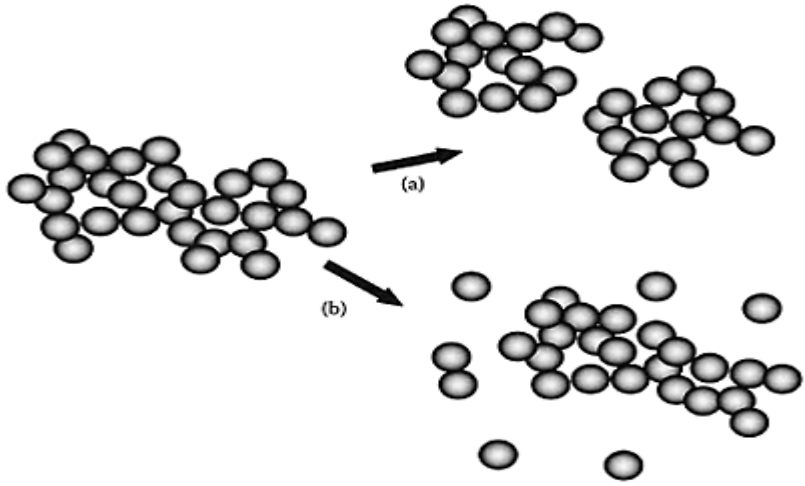

Figure 5.14 Modes of aggregate breakage: (a) Splitting into roughly equal fragments; (b) erosion of small particles.

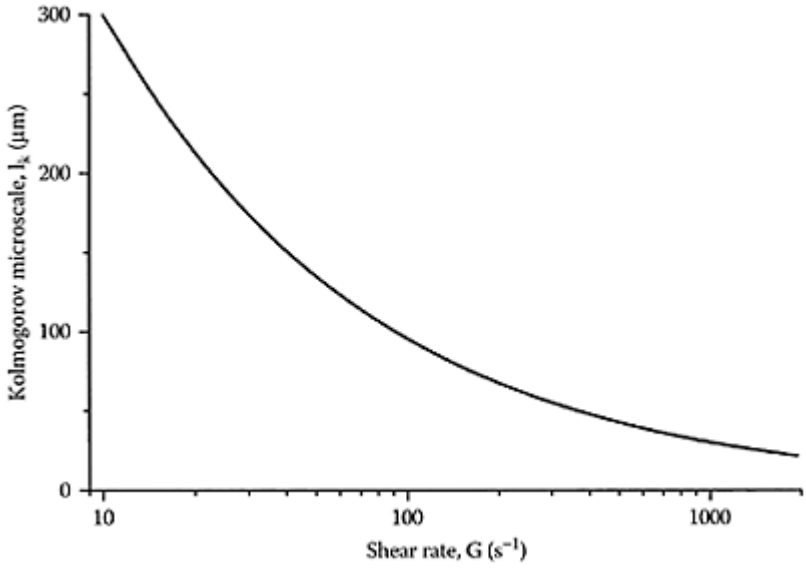

Figure 5.15 Kolmogorov microscale as a function of shear rate.

microscale, l_K, depends on the kinematic viscosity of the fluid, v, and the energy dissipation per unit mass, ε:

$$l_K = \left(\frac{\nu^3}{\varepsilon}\right)^{1/4} \tag{5.39}$$

Values of l_K as a function of mean shear rate, $\bar{G}$(related to energy dissipation rate through Equation 5.26), are shown in Figure 5.15 The fluid is assumed to be water at 25°C. For shear rates in the range of 50–100 s^{-1} (typical of many aggregation processes), the microscale is of the order of 100 μm.

Aggregates of around the same size as the turbulence microscale or smaller are believed to break mainly by surface erosion, whereas larger aggregates tend to undergo splitting. However, there is not a great deal of experimental evidence on this point. As a very approximate rule of thumb, it is sometimes assumed that the maximum aggregate size in turbulent flow is of the same order as the turbulence microscale. However, this ignores the point that aggregates can differ greatly in strength, which must also affect the limiting size. It seems obvious that stronger flocs would have a larger d_{max} value under given conditions.

Thinking in terms of the force, f_0, required to split an aggregate, this must depend on the number of particle-particle bonds broken and the strength of these bonds. The number of bonds broken depends on the cross-sectional area of the broken fragments. It is reasonable to define floc strength, σ, as the ratio of the rupture force to the cross-sectional area:

$$\sigma = \frac{4f_0}{\pi d^2} \tag{5.40}$$

The aggregate strength expressed in this way is a force per unit area, and values from a few Nm^{-2} up to around 1000 Nm^{-2} have been reported, depending on the system. This means that, for typical aggregates of around 50 μm in size, the actual force required for breakage would lie somewhere in a range from nN to μN.

The number of particle-particle bonds per unit area must depend on the density of the aggregate and hence on the fractal dimension. For open, low-density aggregates (low d_F), there will be few bonds per unit area and so the aggregates should be relatively weak. More dense aggregates should have higher strength. It may be that denser aggregates are more prone to surface erosion, rather than splitting, although, again, there is little experimental evidence. Furthermore, as aggregates grow, they become less dense, and the strength, as defined by Equation (5.40), should decrease. This would be another factor determining the maximum aggregate size.

We shall not go more deeply into the subject of aggregate strength and breakage here. This would require a lot of mathematical detail, and there are still considerable difficulties in applying the results to real aggregation processes.

Further reading

Bache, D.H., Floc rupture and turbulence: a framework for analysis, *Chem. Eng. Sci.*, 59, 2521, 2004.

Bushell, G.C., Yan, Y.D., Woodfield, D., Raper, J., and Amal, R., On techniques for the measurement of the mass fractal dimension of aggregates, *Adv. Colloid Interface Sci.*, 95, 1, 2002.

Elimelech, M., Gregory, J., Jia, X., and Williams, R.A., *Particle Deposition and Aggregation. Measurement, Modelling and Simulation*, Butterworth Heinemann, Oxford, 1995.

Mandelbrot, B., *The Fractal Geometry of Nature*, W.H.Freeman, New York, 1982.

Mühle, K., Floc stability in laminar and turbulent flow, in *Coagulation and Flocculation*, Dobias, B., (Ed.), Marcel Dekker, New York, 1993.

Thomas, D.N., Judd, S.J., and Fawcett, N., Flocculation modeling: a review, *Water Research*, 33, 1579, 1999.

chapter six
Coagulation and flocculation

6.1 Terminology

6.1.1 "Coagulation" and "flocculation"

This chapter is concerned with processes in which small particles in water form larger aggregates that can be more easily removed by physical separation processes such as sedimentation, flotation, and filtration. Assuming that the particles are initially stable in the colloidal sense (see Chapter 4), then there are two essential steps in the aggregation process, which are shown schematically in Figure 6.1:

- Destabilization of particles
- Collisions of particles to form aggregates

We shall be dealing mainly with particles that are stable by virtue of their surface charge and hence electrical double-layer repulsion. In that case, destabilization involves either an increase in ionic strength or a neutralization of the particle charge, as discussed in Chapter 4. Adding salts simply to increase ionic strength is almost never a practical option and other additives have to be used, as described later in this chapter. The purpose of the destabilization step is to make the *collision efficiency*, α, as high as possible, ideally with α=1, so that every collision leads to aggregation.

Even when particles are fully destabilized, so that the α=1, collisions are essential if aggregates are to be formed. We saw in Chapter 5 that the particle collision frequency is greatly dependent on the particle concentration and on the collision mechanism. For dilute dispersions, where the collision frequency may be very low, it is possible for particles to be fully destabilized but to show very little aggregation over appreciable time scales. Because of the short-range nature of colloidal interactions, it is usually possible to treat the destabilization and collision processes as independent. In other words,

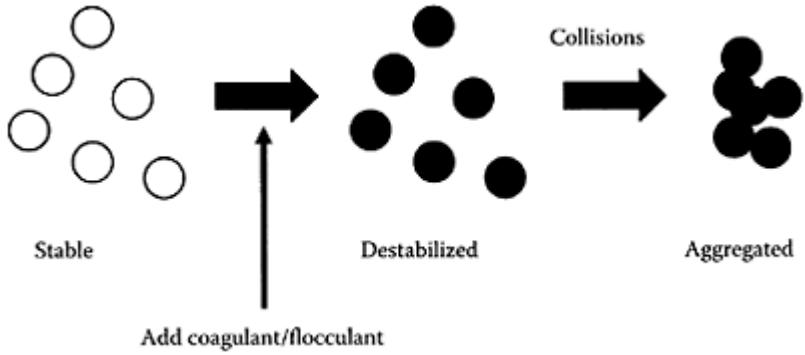

Figure 6.1 Destabilization and aggregation of particles.

it is nearly always safe to assume that the collision frequency is not affected by colloid interactions.

So far in this book, the term "aggregation" has been used in a generic sense, to mean any process whereby particles cluster together to form larger units. It is now time to address the question of other terminology, in particular the widely used terms *coagulation* and *flocculation*. Unfortunately there is no general agreement on how these terms should be used and there are at least two conventions in widespread use.

In the colloid science community it is common to restrict the term *coagulation* to cases where particles are destabilized by simple salts or by charge neutralization and the aggregates *(coagula)* tend to be small and dense. *Flocculation* is then restricted to the cases where polymer bridging is the dominant mechanism and aggregates *(flocs)* tend to be larger and more open in structure. Because of the fractal nature of aggregates (see Chapter 5, Section 5.3.1) it is natural that larger structures tend to be more open and less dense. So the distinction between small, compact coagula and larger, less dense flocs is an inevitable consequence of the stronger interparticle binding in the case of polymers, leading to larger aggregates. Another complication is that *flocculation* is sometimes applied to cases where aggregation occurs in a *secondary minimum* (see Chapter 4, Section 4.4.1).

Another, quite different convention is commonly used in the area of water and wastewater treatment. According to this usage, *coagulation* refers to destabilization, by the dosing of appropriate additives, and *flocculation* implies the formation of aggregates, usually by some form of fluid motion (i.e., orthokinetic aggregation). These correspond to the two stages in Figure 6.1 and could be regarded as chemical and physical aspects of the aggregation process.

6.1.2 Destabilizing agents

Because of the first distinction drawn between coagulation and flocculation, the additives used to cause destabilization of colloids may be called *coagulants* or *flocculants*, depending on their mode of action. Coagulants would then be inorganic salts, including

those containing specifically adsorbing counterions, and flocculants would be long-chain polymers, which give bridging interactions.

Although there are potentially many different kinds of destabilizing agents, the vast majority of those used in practice fall into one of just two categories:

- Hydrolyzing metal coagulants
- Polymeric flocculants

The nature of these materials and their mode of action will be discussed in the following sections.

6.2 Hydrolyzing metal coagulants

The most widely used coagulants are based on aluminum and ferric salts, such as aluminum sulfate ("alum") and ferric chloride. Originally, it was thought that their action was a result of the trivalent nature of the metals, giving Al^{3+} and Fe^{3+} ions in solution, which are expected to be very effective in destabilizing negatively charged colloids. However, this is a greatly over-simplified view because trivalent metal ions are readily *hydrolyzed* in water, which has an enormous effect on their behavior as coagulants.

6.2.1 Hydrolysis of metal cations

In some cases, metal ions in water exist mainly in the form of simple hydrated cations. This is the case for alkali metal ions such as Na^+ and K^+. Because of the polar nature of water, such cations are *hydrated* to some extent, which means they are surrounded by a certain number of water molecules held by electrostatic attraction between the positive metal ion and the negative (oxygen) ends of the water molecules. It is reasonable to think in terms of a *primary hydration shell*, where water molecules are in direct contact with the central metal ion and more loosely held water in a secondary hydration shell.

In the case of the trivalent metal ions Al^{3+} and Fe^{3+} it is known that the primary hydration shell consists of six water molecules in octahedral coordination, as shown in Figure 6.2 (a). Now, because of the high positive charge on the central metal ion, there is a tendency for electrons to be drawn toward the metal from the water molecules, and this can lead to the dissociation of a proton, H^+, leaving a hydroxyl group attached and a reduced positive charge for the metal, as shown in Figure 6.2 (b). Because the process essentially involves the splitting of water molecules, it is known as *hydrolysis*. Because hydrolysis causes the release of a hydrogen ion into solution, it is greatly dependent on the pH. High pH values promote dissociation and vice versa. Furthermore, as each proton is released, the decreasing positive charge makes further dissociation more difficult. It follows that with increasing pH there is a sequence of hydrolysis equilibria, which can be written as follows:

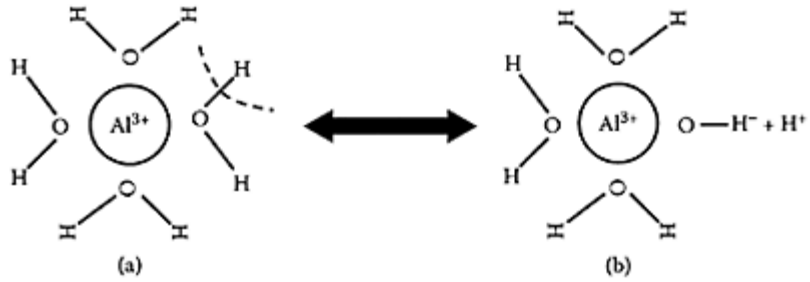

Figure 6.2 Hydrolysis of Al^{3+}. (a) Hydrated aluminum cation. (Note: only 4 of 6 water molecules shown.) (b) After loss of H^+ to give $Al(OH)^{2+}$.

$$Me^{3+} \rightarrow Me(OH)^{2+} \rightarrow Me(OH)_2^{+} \rightarrow Me(OH)_3 \rightarrow Me(OH)_4^{-}$$

(For simplicity, water molecules in the hydration shell are omitted.)

Each of the stages in the hydrolysis process has an appropriate equilibrium constant:

$$M^{3+} + H_2O \leftrightarrow M(OH)^{2+} + H^+ \qquad K_1$$

$$M(OH)^{2+} + H_2O \leftrightarrow M(OH)_2^{+} + H^+ \qquad K_2$$

$$M(OH)_2^{+} + H_2O \leftrightarrow M(OH)_3 + H^+ \qquad K_3$$

$$M(OH)_3 + H_2O \leftrightarrow M(OH)_4^{-} + H^+ \qquad K_4$$

These are defined in the conventional manner, so that in the case of K_2, for example:

$$K_2 = \frac{[M(OH)_2^+][H^+]}{[M(OH)^{2+}]} \tag{6.1}$$

where square brackets indicate molar concentrations of the various species.

For Al and Fe(III), the uncharged hydroxide, $M(OH)_3$, has very low solubility in water and is likely to form a precipitate over a certain range of pH. This precipitation is of great importance in the action of hydrolyzing metal coagulants (see later). As well as the equilibrium constants listed earlier, a solubility constant for the metal hydroxide is also needed, based on the following dissolution of the solid phase, $M(OH)_3(s)$:

$$M(OH)_3(s) \leftrightarrow M^{3+} + 3OH^- \qquad K_S$$

$$K_S = [M^{3+}][OH^-]^3 \tag{6.2}$$

Table 6.1 Equilibrium constants (pK values) for Al and Fe(III) hydrolysis and solubility of amorphous hydroxides (values for 25°C and zero ionic strength)

	pK_1	pK_2	pK_3	pK_4	pK_{Sam}
Al^{3+}	4.95	5.6	6.7	5.6	31.5
Fe^{3+}	2.2	3.5	6	10	38

If true equilibrium were attained, then the appropriate solubility constants would be those for the stable crystalline forms such as gibbsite and goethite, in the case of Al and Fe, respectively. However, these are usually formed slowly (usually weeks or months). From the standpoint of coagulation processes it is much more relevant to consider the solubility constants *(K_{Sam})* of the amorphous precipitates that form initially. Unfortunately these values are subject to some uncertainty and only estimated values can be given. They are usually at least 100-fold larger than values for the corresponding crystalline solids, so that the amorphous material is effectively much more soluble.

Table 6.1 gives some values for the hydrolysis and solubility constants for Al and Fe(III) species in water at 25°C and at zero ionic strength, so they are appropriate for low salt concentrations, typical of many natural waters. The constants are given in the conventional p*K* form, where $pK = -\log_{10} K$.

Using these p*K* values, it is possible to calculate, as a function of pH, the concentrations of the various dissolved hydrolysis products in equilibrium with the amorphous hydroxide precipitate. Because of uncertainties over the solubility constants for the amorphous precipitates, the results may not be completely reliable, but they give a useful indication of the relative importance of the different species over a range of pH values. Figure 6.3 is a *speciation* diagram showing the results of such calculations for Al and Fe(III), based on the values in Table 6.1. The total concentration of dissolved species in equilibrium with the amorphous precipitate is effectively the solubility of the metal at a given pH value. It is evident from Figure 6.3 that there is a minimum solubility that occurs around neutral pH for both metals. Note that the minimum solubility of Fe(III) is much lower than that of Al and that the minimum is considerably broader. It is also apparent that, in the case of Al, the anionic form $Al(OH)_4^-$ (aluminate) is the dominant dissolved species above neutral pH.

Another way of showing the speciation data is to plot the mole fraction of each species in relation to the total dissolved amount in equilibrium with the amorphous hydroxide. This has been done in Figure 6.4 for Al and Fe(III). These results show considerable differences between the two metals. In the case of Al, the predominant species are the trivalent ion, Al^{3+}, at low pH (up to about 4.5) and the aluminate ion, $Al(OH)_4^-$, at pH values higher than about 7. The intermediate species make only minor contributions at pH values in the region of about 4–6.5. For Fe(III) the various species are spread over a much broader pH range (about 8 units) and each hydrolysis product

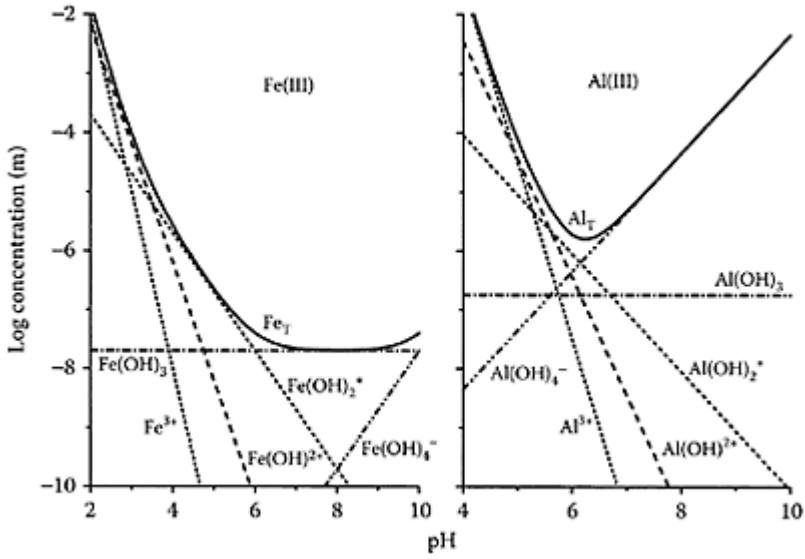

Figure 6.3 Speciation diagrams for Fe(III) and Al(III). (Note: only monomeric hydrolysis products shown.)

is dominant at some pH values. This is the expected behavior for hydrolysis of metal ions. The reason that Al species are "squeezed" into a much narrower pH range is believed to be the result of a transition from octahedral coordination in $Al^{3+}{\cdot}6H_2O$ to the tetrahedral $Al(OH)_4^-$. In the case of Fe(III), octahedral coordination is maintained throughout. It is also worth noting that the *soluble*, uncharged $Fe(OH)_3$ is the predominant dissolved Fe species in the pH range 7–9 (although the actual concentration is only around 2×10^{-8} M). The corresponding Al species, $Al(OH)_3$, is always a minor dissolved component in *relative* terms, although it is at least 10 times more soluble than $Fe(OH)_3$.

Our discussion so far has ignored certain complications, one of which is the effect of various anions that can influence hydrolysis equilibria. For instance, it is known that fluoride forms strong complexes with Al and this gives a greater aluminum solubility than would be predicted on the basis of the results in Figure 6.3 Another point is that only *monomeric* hydrolysis products have been considered, whereas, under some conditions, *polynuclear* species can be important. These form the subject of the next section.

6.2.2 Polynuclear hydrolysis products

In addition to the monomeric hydrolysis products considered earlier, there are many possible *polynuclear* forms that may be important. For Al these include $Al_2(OH)_2^{4+}$ and $Al_3(OH)_4^{5+}$, and there are similar forms for Fe(III). However, these are not likely to be

significant at the low concentrations of the metals usually used in coagulation. In practice, only the monomeric forms and the hydroxide precipitate are likely to be important.

Polynuclear hydrolysis products can be prepared under certain conditions. The best known of these is $Al_{13}O_4(OH)_{24}^{7+}$ or "Al_{13}," which can be formed by controlled neutralization of aluminum salt solutions or by several other methods. This tridecamer has the so-called keggin structure, consisting of a central tetrahedral AlO_4^{5-} unit surrounded by 12 Al octahedra with shared edges. The tetrahedral and octahedral Al sites can be easily distinguished in the ^{27}Al NMR spectrum. Under appropriate conditions, Al_{13} forms fairly rapidly and essentially irreversibly remaining stable in aqueous solutions for long periods. The tridecamer is believed to be present fairly widely in the natural aquatic environment, such as in acid forest soil water.

Other polynuclear species, such as the octamer, $A1_8(OH)_{20}^{4+}$, have been proposed, based on coagulation data. However, there is no direct evidence for the octamer and it is unlikely to be significant in practice.

There are many commercial products based on prehydrolyzed metal salts. In the case of aluminum, a common example is the class of materials known as *polyaluminum chloride* (PACl), which can be produced by controlled neutralization of aluminum chloride solutions. It is likely that many of these products contain substantial amounts of the tridecamer Al_{13}. In the case of aluminum sulfate it is difficult to prepare prehydrolyzed forms with high degrees of neutralization because sulfate encourages hydroxide precipitation. The presence of small amounts of dissolved silica can significantly improve the stability, and the resulting product is known as *polyaluminosilicate-sulfate* (PASS).

There are corresponding products containing polymerized iron species, although these are not as widely used as PACl.

6.2.3 Action of hydrolyzing coagulants

There are essentially two important ways in which hydrolyzing coagulants can act to destabilize and coagulate negatively charged colloids. At low concentrations and under suitable pH conditions, cationic hydrolysis products can adsorb and neutralize the particle charge, hence causing destabilization and coagulation. At higher doses of coagulant hydroxide precipitation occurs and this plays a very important role—giving the so-called sweep coagulation or sweep flocculation.

6.2.4 Charge neutralization by adsorbed species

At very low concentrations of metal, only soluble species are present—the hydrated metal ion and various hydrolyzed species (see Figure 6.3). It is generally thought that hydrolyzed cationic species such as $Al(OH)^{2+}$ are more strongly adsorbed on negative surfaces than the free metal ion and so can effectively neutralize surface charge. Generally, charge neutralization with aluminum salts occurs at low metal concentrations, usually of the order of

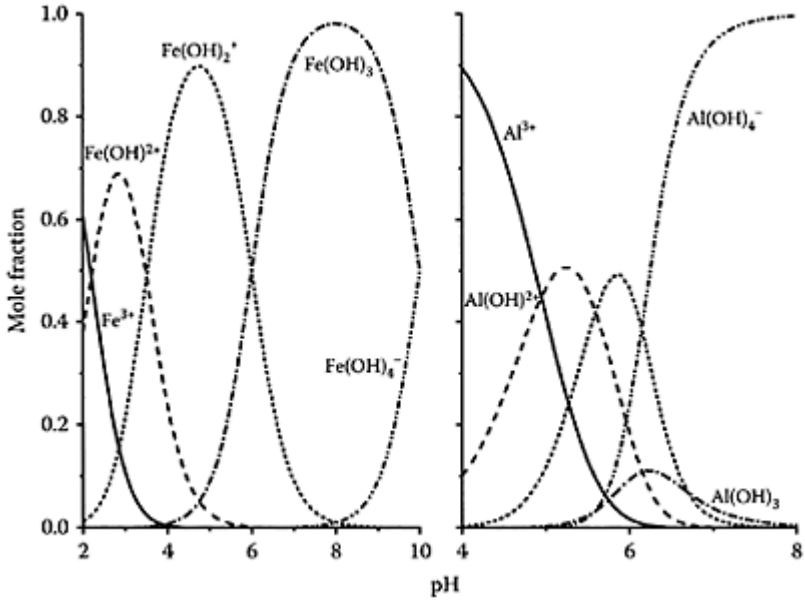

Figure 6.4 Proportion (mole fraction) of hydrolyzed Fe(III) and Al(III) species relative to total soluble metal concentration.

a few micromoles/L at around neutral pH. It has been found that, for several inorganic suspensions at pH 6, the amount of Al needed to neutralize the surface charge is around 5 μmoles per m^2 of particle surface (of the order of 130 μg Al per m^2). However, even at very low metal concentrations, the solubility of the amorphous hydroxide may be exceeded. Also, in the region of neutral pH, cationic hydrolysis products represent only a small fraction of the total soluble metal, especially for Al (Figure 6.4). The fact that charge neutralization is commonly observed in such cases suggests that the effective species might be colloidal hydroxide particles. In the case of aluminum hydroxide, the point of zero charge (pzc) (see Chapter 3, Section 3.1.2), is around pH 8, so the precipitate particles should be positively charged at lower pH values. For ferric hydroxide the pzc is somewhat lower, around pH 7. Even when the bulk solubility is not exceeded, it is possible that some form of *surface precipitation* may occur as a result of nucleation at the surface.

Actually, it is difficult to distinguish between surface precipitation and the attachment of colloidal hydroxide particles that have been precipitated in solution. A combination of these effects may be most likely in practice and forms the basis of the *precipitation charge neutralization* (PCN) model, which is illustrated schematically in Figure 6.5.

Whatever the precise nature of the charge-neutralizing species, they are likely to be capable of charge reversal at higher dosages. This means that there will be a characteristic *optimum dosage* at which coagulation is most effective. At higher dosages,

particles become positively charged and *restabilized*. As discussed in Chapter 4, Section 4.4.3, the optimum dosage must

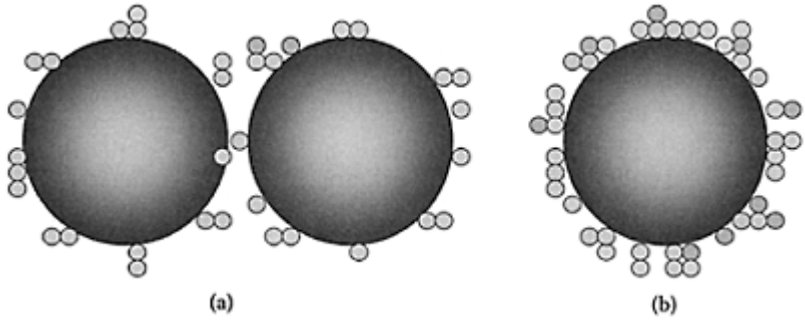

Figure 6.5 Precipitation charge neutralization (PCN) model, showing (a) charge neutralization and (b) charge reversal (restabilization) of particles by precipitated hydroxide colloids. (After Dentel, 1991.)

depend on the particle concentration, but in practice the value is often low. Sometimes, the optimum dosage range can be narrow, so precise dosing control is necessary.

Another disadvantage of relying on charge neutralization is that, for low particle concentrations, the collision rate and hence the aggregation rate will be low, and long times may be needed to give sufficiently large aggregates (flocs). Neutralizing surface charge by small adsorbed species does nothing to enhance the collision rate, although, of course, the collision *efficiency* can be greatly enhanced.

Some of the advantages claimed for prehydrolyzed coagulants are supposed to be a result of the presence of highly charged cationic species, such as $Al_{13}O_4(OH)_24^{7+}$. The fact that this ion carries 7 positive charges suggests that it would be very strongly adsorbed on negative surfaces and would be effective in neutralizing particle charge. (It should be noted that the Al_{13} species has only about half an elementary charge per Al atom, whereas forms such as Al^{3+} and $Al(OH)^{2+}$ can, in principle, deliver more charge per Al.) Accepting that species such as Al_{13} can be more effective in neutralizing charge, it is still difficult to see how, at the optimum dosage, the coagulation rate could be significantly higher than with other adsorbing cationic species.

6.2.5 "Sweep" flocculation

In most practical water treatment operations, metal coagulants are added at dosages much higher than the solubility of the amorphous hydroxide and extensive precipitation occurs. For reasons that are still not fully understood, this can give much more effective separation than simple charge neutralization. The most likely explanation is that the original impurity particles are somehow incorporated into the growing hydroxide

precipitate and are thereby removed from suspension. This enmeshment of particles is generally thought of as a "sweeping" action—hence the terms "sweep coagulation" or "sweep flocculation."

The choice of term is somewhat arbitrary, in view of the terminology discussed in Section 1.1. The hydroxide precipitate could be regarded as "bridging" particles together and hence "sweep *flocculation*" might be the more appropriate term from one point of view. Also, in water treatment, the formation of large hydroxide aggregates requires some form of agitation, so that orthokinetic collisions are important, and this again supports the use of the term "flocculation." However "sweep coagulation" is also widely used and it might be better to regard these terms as interchangeable. The aggregates formed as a result of hydroxide precipitation are almost universally known as "flocs." It is confusing that the additives used are mostly known as "coagulants."

Sweep flocculation almost always leads to faster aggregation than charge neutralization and gives stronger and larger flocs. The reason for the higher aggregation rate is not hard to find. The production of a hydroxide precipitate gives a big increase in the effective particle concentration and hence a greater collision rate, according to Smoluchowski theory (Chapter 5). Hydroxide precipitates are formed from large numbers of colloidal particles, which form very soon after dosing. The aggregation of these small particles gives low-density flocs, with a relatively large volume. According to the theory of orthokinetic aggregation (Equation 5.24), the rate is directly proportional to the volume fraction of suspended particles, and this can be vastly increased by hydroxide precipitation. This is the main reason why sweep flocculation is so much more effective than charge neutralization. The flocs produced under "sweep" conditions are also stronger and therefore grow larger for the same shear conditions. However, hydroxide flocs are still weak compared to those formed by polymeric flocculants (see Section 6.3).

A major advantage of sweep flocculation is that it does not much depend on the nature of the impurity particles to be removed, whether bacteria, clays, oxides, or others. For relatively dilute suspensions, the optimum coagulant dosage is that which gives the most rapid hydroxide precipitation and is practically independent of the nature and concentration of suspended particles.

The large volume associated with hydroxide flocs leads to a significant practical problem—the production of large amounts of sludge that needs to be disposed of in some way. In a typical water treatment plant most of the sludge produced is associated with metal hydroxide rather than the impurities removed from water. Although there is usually no significant restabilization in the case of sweep flocculation and hence no sharp optimum dosage region, overdosing is best avoided to restrict the volume of sludge produced.

The action of prehydrolyzed coagulants, such as polyaluminum chloride, at typical dosages also very likely involves hydroxide precipitation and sweep flocculation, although this point has not been thoroughly investigated. There is evidence that the nature of the precipitate differs from that formed with "alum."

6.2.6 Overview

With increasing dosage of hydrolyzing coagulant to a suspension of negatively charged particles, four distinct zones are recognized:

Zone 1: Very low dosage; particles still negative and hence stable

Zone 2: Dosage sufficient to give charge neutralization and hence coagulation

Zone 3: Higher dosage giving charge reversal and restabilization

Zone 4: Still higher dosage giving hydroxide precipitation and sweep flocculation

Figure 6.6 shows the results of a standard jar *test* procedure, usually used in water treatment applications. In this procedure, a suspension is dosed with different amounts of coagulant under standard mixing and sedimentation conditions. Usually there is a brief *rapid mix* period immediately after dosing. This is followed by a longer period of *slow stirring* during which flocs may be formed as a result of orthokinetic aggregation. These flocs are then allowed to settle for a standard period, after which a sample of the supernatant water is taken and its turbidity is measured. This *residual turbidity* gives a good indication of the degree of removal during sedimentation and hence of the effectiveness of the coagulation/flocculation process.

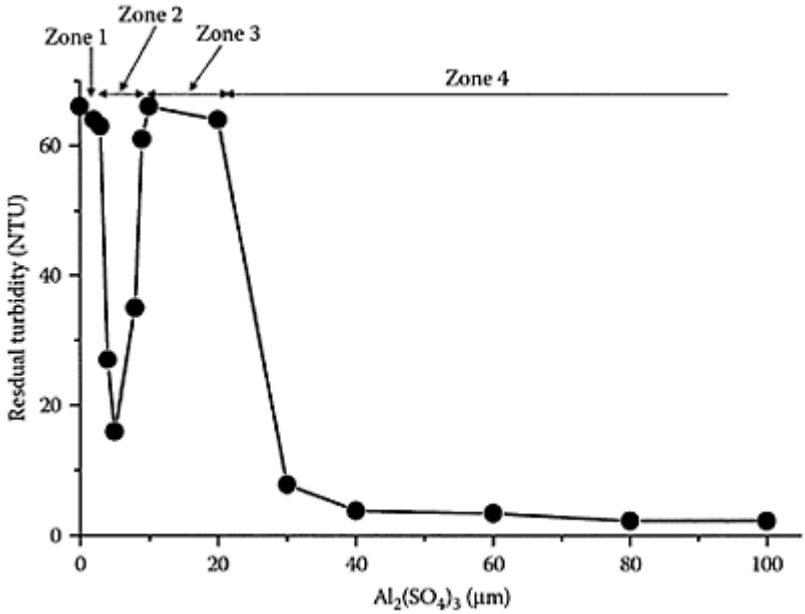

Figure 6.6 Residual turbidity of kaolin suspensions after coagulation with aluminum sulfate over a range of concentrations at pH 7. (Replotted from data of J.Duan, Ph.D. Thesis, University of London, 1997.)

Figure 6.6 shows that, at very low coagulant dose, the residual turbidity is high, indicating little or no sedimentation (Zone 1). As the dose is increased, there is a fairly narrow range (Zone 2) where there is a significant reduction in residual turbidity. This is the region of charge neutralization by adsorbed species, and it is very often found that the particle charge (as measured, for instance, by electrophoretic mobility or streaming current; see Chapter 3) is around zero. At higher dosages residual turbidity is again high, indicating restabilization of the particles as a result of excess adsorption and charge reversal (Zone 3). Finally, at still higher dosages, there is a substantial reduction in residual turbidity because of hydroxide precipitation and "sweep" flocculation (Zone 4). Note that the residual turbidity in Zone 4 is lower than that in Zone 2, showing that sweep flocculation gives larger, faster-settling flocs than those formed by charge neutralization. Also, as mentioned earlier, there is no restabilization after Zone 4.

The behavior shown in Figure 6.6 is typical of aluminum salts at around pH 7. Under these conditions the hydroxide precipitate is positively charged. At pH values near to the isoelectric point (around pH 8) Zone 2 may not be apparent and only sweep flocculation is operative.

6.2.7 Practical aspects

There are several important factors that can greatly affect the performance of hydrolyzing coagulants. These include the effects of various anions and the influence of temperature.

Several common anions can form complexes with aluminum and iron (III) and can significantly affect hydroxide precipitation. An important example is sulfate, which is naturally present in water and may be added in the form of aluminum or ferric sulfate in water treatment. Sulfate coordinates moderately strongly with Al, but the main effect is on the precipitation process. On the positive side of the isoelectric point of aluminum hydroxide (i.e., below about pH 8) sulfate can adsorb on the precipitate and reduce its positive charge. This means that the colloidal precipitate can aggregate more rapidly to give large hydroxide flocs.

Temperature has effects that are important in practice. In particular, at rather low temperatures, conventional aluminum coagulants tend to perform less well for various reasons. Some prehydrolyzed coagulants appear to be less affected by low temperatures and are often preferred for applications in cold regions.

Another advantage of prehydrolyzed coagulants such as polyaluminum chloride is that, at effective dosages, they produce less sludge than simple metal salts. This may be partly because they can be effective at lower concentrations.

Hydroxide flocs, as formed during sweep flocculation, tend to be weak and are easily disrupted under high shear conditions. Furthermore, the breakage can be irreversible to some extent so that flocs do not easily reform when the shear rate is reduced. Polymeric flocculants, either alone or in combination with hydrolyzing metal salts, can give significantly stronger flocs. These additives will be considered in the next section.

6.3 Polymeric flocculants

We have seen in Chapter 4 that adsorbed polymers may give repulsion *(steric repulsion)* or attraction *(polymer bridging)* between particles. This section is concerned with the destabilizing action of polymers. For adsorbing nonionic polymers, attraction between particles is entirely the result of the "bridging" effect. However, for charged polymers (or *polyelectrolytes*), there is also the possibility that *charge neutralization* can play a role. Although both of these effects may operate simultaneously, it is convenient to treat them separately. We shall first consider the nature of polymeric flocculants and their adsorption on particles in water.

6.3.1 Nature of polymers and polyelectrolytes in solution

Polymers are long-chain molecules consisting of at least one type of repeating unit (or *monomer*). They may vary in molecular weight from a few thousand up to many millions (or up to many thousand monomer units). Polymers may be essentially *linear* in nature or have extensive branching. However, nearly all effective polymeric flocculants have a linear structure, and we shall only consider this type.

If fully extended, a polymer with a very high molecular weight could have a length approaching 100 μm (0.1 mm). However, in solution, polymers adopt a *random coil* configuration with much smaller dimensions (usually less than 1 μm). The conformation of a polymer random coil can be considered in terms of a *random walk*, analogous to the treatment of brownian motion (Chapter 2, Section 2.3.2). It follows that effective size of a polymer molecule in solution (e.g., the radius of gyration) is proportional to the square root of the molecular weight.

If the monomer units have ionizable groups, they can become charged. This may lead to significant repulsion between segments of the polymer chain and hence an expansion from the typical random coil configuration. Repulsion between charged groups can be "screened" by ions in solution, in much the same way that the range of double-layer repulsion is reduced at high salt concentration (see Chapter 4). For this reason, ionic strength has an important effect on polyelectrolyte chain expansion in aqueous solutions. These concepts are illustrated schematically in Figure 6.7.

Characterization of polymers in solution is most often carried out by *light scattering* or *viscometry*, both of which depend on the relatively large size of polymer molecules. Both of these methods can, in principle, give information on molecular weight, although the results are not always easy to interpret. In the case of viscometry, it is convenient to think in terms of an *intrinsic viscosity*, which is derived as described later.

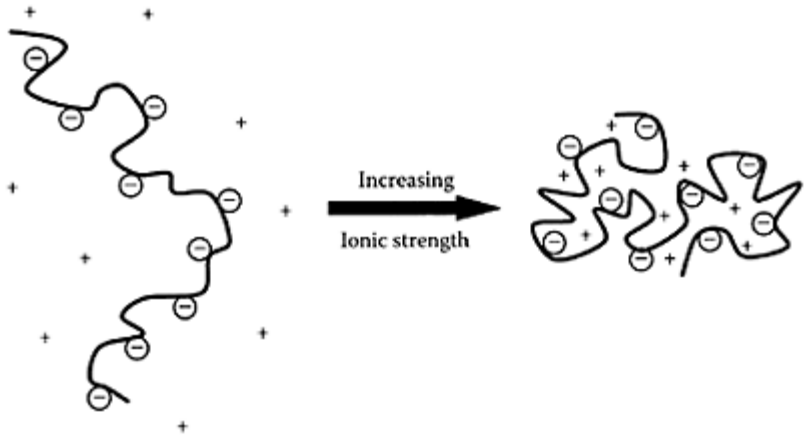

Figure 6.7 Showing the effect of ionic strength on the conformation of an anionic polyelectrolyte molecule in solution. Higher salt concentrations cause the chain to adopt a random coil arrangement. At low salt concentration the chain is more extended.

Viscosity of polymer solutions can be derived most easily by a capillary flow method. The time required for a fixed volume of solution to pass through a capillary tube under defined conditions is directly proportional to the viscosity. Hence the viscosity of a polymer solution, relative to the solvent (water) is simply the ratio of the flow times. The *specific viscosity* of a solution is defined by the following:

$$\mu_{sp} = \frac{\mu - \mu_0}{\mu_0} = \frac{t - t_0}{t_0} \tag{6.3}$$

where μ and μ_0 are the viscosities of the polymer solution and water, respectively, and t and t_0 are the corresponding capillary flow times.

The *reduced viscosity* is just the specific viscosity divided by the polymer concentration, *c:*

$$\mu_{red} = \mu_{sp}/c \tag{6.4}$$

If the specific viscosity were proportional to concentration, then the reduced viscosity should be constant, independent of concentration. This is not usually observed, but a plot of reduced viscosity against concentration is often nearly linear, and it is possible to extrapolate the line to zero concentration, which gives the *intrinsic viscosity, [μ]*. From the definition of *[μ],* it follows that this quantity has dimensions of reciprocal

concentration and depends on the concentration units used. For instance, polymer concentrations might be expressed as g/L, and intrinsic viscosity would have units of L/g.

For polyelectrolytes it is usual to determine intrinsic viscosity at high salt concentrations (typically 3M NaCl) to minimize charge effects on polymer conformation, so that the chains adopt a random coil configuration.

The intrinsic viscosity is related to polymer molecular weight by the *Mark-Houwink* equation:

$$[\mu] = kM_V^{\ a} \tag{6.5}$$

where k and a are empirical constants, which have to be determined experimentally for each polymer. For many polymers the exponent a has a value of around 2/3 and k is of the order of $1–10\times10^{-5}$ L/g. The subscript in M_V indicates that this is a *viscosity-average* molecular weight. There will generally be a molecular-weight distribution for a given polymer, and the kind of average derived depends on the experimental technique used.

Because the conversion of intrinsic viscosity to molecular weight is subject to some uncertainty, manufacturers of polymeric flocculants sometimes give values of *[μ]* rather than molecular weight.

In the case of polyelectrolytes another important characteristic is the *charge density*, which can be conveniently measured by the technique of *colloid titration*. Charge density is properly expressed as chemical equivalents of charged groups per unit mass of polymer (usually in meq/g). However, for many commercial products (usually copolymers—see Section 6.3.2), manufacturers may give information in terms of the mole fraction of the ionic component. So, if a cationic polyelectrolyte is stated to be “30% charged,” it usually means that 30% of the monomer units are cationic and 70% are nonionic.

6.3.2 Examples of polymeric flocculants

There are many types of polymeric flocculant used in various solid-liquid separation processes, although those in common use are restricted to relatively few categories. Broadly, these flocculants may be classified into three types, depending on their charge:

- Nonionic (uncharged)
- Anionic (negatively charged)
- Cationic (positively charged)

Within these categories there can be many different chemical structures, very wide ranges of molecular weight, and great differences in charge density.

There are several examples of polymeric flocculants that are based on natural products. These include isinglass, gelatine, starch, and alginates, which are still used, especially in the clarification of potable liquids such as beer and wine. A modified form of starch *(cationic starch)* is commonly used

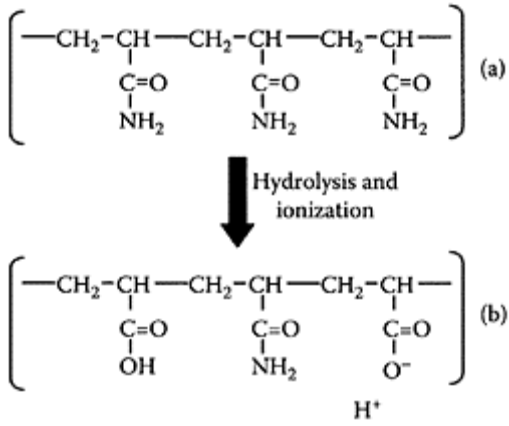

Figure 6.8 (a) Segment of a polyacrylamide molecule. (b) Same segment after some hydrolysis of amide groups and ionization of resulting carboxylic acid groups.

in the papermaking industry. *Chitosan* is a cationic polymer derived from shells, especially crab shells, and is often used as a flocculant in water treatment, biotechnology, and other fields. Extracts from the crushed nuts and seeds of certain plants have been used for centuries to clarify turbid waters. Starting materials include the seeds of the *Nirmali* tree in India and of the *Moringa oleifera* plant in Sudan, Peru, and Indonesia (where it is known as *kelor*).

However, the vast majority of polymeric flocculants in common use are entirely synthetic in origin. Many of these are based on acrylamide monomer, which can be easily polymerized to give products of very high molecular weight (up to 20 million or more). The structure of polyacrylamide (PAM) is shown in Figure 6.8. The first form shown has no ionic groups and so would be a nonionic polymer. However, the amide groups attached to the polymer backbone may be hydrolyzed, giving *carboxylic acid* groups, as shown. (The modified monomer unit then becomes like those in polyacrylic acid.) At pH values of around 5 and higher the acid groups ionize to give *carboxylate* ions, so that the polymer acquires anionic character.

Actually, some inadvertent hydrolysis of polyacrylamide usually occurs during manufacture, so that even nominally “nonionic” PAM can be slightly anionic in nature. Controlled hydrolysis can give moderate anionic charge, but it is also possible to produce anionic PAM by copolymerization of acrylamide and acrylic acid. In this way up to 100% anionic charge can be achieved (in which case the polymer would be pure polyacrylic acid). Charge densities of anionic PAMs are usually expressed in percentage terms. For instance, “30% charge” implies that 30% of the acrylamide groups are replaced by acrylic acid units.

Cationic polyelectrolytes based on polyacrylamide are also widely used. These can be prepared by reaction of the amide groups with dimethylamine and formaldehyde

(Mannich reaction) or by copolymerization of acrylamide with a cationic monomer, such as dimethyl aminoethyl methacrylate (DMAEMA). In both cases, the resulting amine groups may be quaternized to give strongly cationic groups (i.e., where the charge is largely independent of pH). Again, the charge density is often expressed in terms of the mole fraction of cationic groups. Cationic polyelectrolytes based on polyacrylamide can be prepared with very high molecular weights.

Other examples of commercial polymeric flocculants are s follows:

Nonionic: Polyvinyl alcohol (PVA), polyethylene oxide (PEO)
Anionic: Sodium polystyrene sulfonate
Cationic: Polyethyleneimine, polydiallyldimethylammonium chloride (polyDADMAC), poly (2-vinyl imidazolene)

6.3.3 Polymer adsorption

For effective flocculation, polymers need to be adsorbed on particles. In general the adsorption of a polymer chain on a surface leads to a more restricted configuration than that of a random coil in solution, so there is some loss of entropy. For this reason, there must be favorable interaction between polymer segments and surface sites. However, the attraction between an individual segment and the surface can be weak because attachment of a polymer chain at many points gives strong interaction overall. In fact, polymer adsorption is often so strong that the process is effectively irreversible.

Adsorption interaction can be of several types, including the following:

- Electrostatic interaction: When a polyelectrolyte adsorbs on a surface with opposite charge—for instance, cationic polymers on negative surfaces—a dominant contribution to the interaction comes from electrostatic attraction, which usually gives very strong adsorption. Even for low molecular weights, adsorption can be essentially quantitative (i.e., all of the added polymer is adsorbed, at least up to the point of charge neutralization). This strong adsorption is an important factor in many practical applications of polymeric flocculants. However, salt effects can be important. High salt concentrations can screen the electrical interaction, giving weaker adsorption.
- Hydrogen bonding: For instance, in the case of polyacrylamide and metal oxides, H-bonds can form between the amide groups of PAM and surface hydroxyl groups.
- Hydrophobic interaction: This is responsible for the adsorption of nonpolar segments of polymer chains to hydrophobic regions of a surface. An example is the adsorption of polyethylene oxide on fine coal particles.
- Ion binding: It is possible (and common) for polyelectrolytes to adsorb on surfaces of like sign (i.e., against an electrostatic repulsion). A common example is the adsorption of anionic PAM on negative surfaces, which in some cases occurs through the intervention of certain metal ions, especially divalent ions such as calcium. These

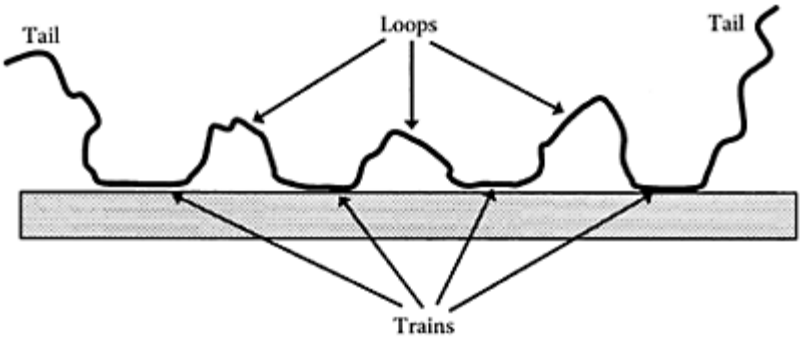

Figure 6.9 Equilibrium conformation of an adsorbed polymer chain (see text).

can link anionic sites on the PAM and surface by the mechanism of "calcium bridging," which is well-known in biology. Some negatively charged particles can be easily flocculated by anionic PAM in the presence of fairly low calcium concentrations (around 1 mM or higher), but in the absence of calcium no flocculation occurs.

When a polymer chain adsorbs on a surface from solution, the original random coil configuration is not maintained and the polymer becomes attached at various points because of one or other of the interactions outlined earlier. Essentially, the chain becomes uncoiled and eventually adopts an equilibrium adsorbed configuration that may be like that shown schematically in Figure 6.9. Polymer segments in an adsorbed chain may be found in the following situations:

- Attached to the surface in *trains*
- Projecting into the solution as *tails* (two per chain)
- In the form of *loops*, linking trains together

It is important to note that the conventional picture in Figure 6.9 represents an *equilibrium* condition, which may take some time to achieve from the instant of first contact of the chain with the surface. Dynamic aspects of polymer adsorption are still not well understood but it is reasonable to assume that, for long-chain polymers, times of several seconds or more could be required for the equilibrium conformation to be achieved. This can be very important in the kinetics of flocculation by polymers (see Section 6.3.6).

The extent of the tails and loops, and hence the effective thickness of an adsorbed polymer layer, depends greatly on the nature of the interaction between polymer segments and surface. In the case of electrostatic attraction between, say, a high-charge cationic polyelectrolyte and a negative surface, it is likely that the chain will adopt a flat adsorbed conformation. However, for weaker interactions, segments of an adsorbed chain will extend further into solution.

6.3.4 Bridging flocculation

When an adsorbing polymer is added to a suspension of particles, a single chain may become attached to two or more particles. This is unlikely to occur simultaneously, except in concentrated suspensions, but once a chain is attached to one particle, collisions with other particles can give *polymer bridging* between them. This has already been discussed briefly in Chapter 4, Section 4.5.4, and illustrated schematically in Figure 4.13.

An essential requirement for bridging flocculation is that there should be sufficient uncoated surface on a given particle for attachment of segments of polymer chains adsorbed on other particles. In other words, the adsorbed amount should not be too high. Also, to give effective binding between particles, the adsorbed amount should not be too low, otherwise insufficient "bridges" would be formed between particles. These considerations lead directly to the idea of an optimum dosage of polymer to give the most effective flocculation. An early idea, introduced by Victor La Mer, was that the optimum dosage corresponds to "half-surface coverage." If the fraction of particle surface covered by polymer is θ, then the uncovered fraction is $1–\theta$ and it is not difficult to show that the fraction of successful collisions (i.e., those between coated and uncoated surfaces) is proportional to the term $\theta(1–\theta)$. This term has a maximum value when $\theta=0.5$, in line with La Mer's "half-surface coverage" idea. However, for adsorbed polymers, it is difficult to define "surface coverage" precisely. Optimum flocculation usually occurs at well below saturation (monolayer) coverage.

It is well established that bridging flocculation gives aggregates (flocs) that are much stronger than those formed when salts are used to destabilize a suspension. Linkage by polymer chains gives stronger bonds between particles than van der Waals forces. The *flexibility* of polymer bridges may also be an important factor in practice. Some extension is necessary before breakage occurs. Because polymer bridging produces strong flocs, they can grow quite large, and because flocs are fractal objects, this means that floc density can become low (see Chapter 5, Section 5.3.3). The commonly observed open structure of polymer-generated flocs can probably be explained in this way.

Although polymer bridging produces strong flocs, which can withstand high shear, when the flocs *are* broken they may not readily reform. This irreversible floc breakage may be a result of rupture of polymer chains, but this aspect is not well understood.

The effectiveness of polymers as bridging flocculants depends greatly on their properties. The most important property is *molecular weight because* this directly influences the effective size of the molecules in solution and in the adsorbed state. Higher-molecular-weight polymers are found to be more effective flocculants. For a similar reason, it is generally found that linear polymers are more effective than branched or cross-linked structures for a given molecular weight.

For polyelectrolytes, the *charge density* can also be an important variable. Highly charged chains will tend to adopt a more expanded configuration (see Figure 6.7) and might be expected to be more effective bridging agents for this reason. However, for like signs of charge (as with anionic PAM and negative particles), highly charged polyelectrolytes will be less readily adsorbed and so might be less effective flocculants. It follows that there may be an *optimum charge density* for most effective flocculation. For PAM flocculants it is sometimes found that there is an optimum degree of hydrolysis (and hence of anionic character). This has been quoted to be in the region of 30% hydrolysis, but it is difficult to find consistent results on this point.

Ionic strength would also be expected to play a significant part in the bridging action of polyelectrolytes. Increasing salt concentration would reduce chain expansion but increase the adsorption on surfaces of like charge. For these reasons, there might be an optimum ionic strength, although there is no convincing evidence in support of this conjecture.

In summary, the important points for bridging flocculation are as follows:

- Higher-molecular-weight polymers are more effective flocculants than low-molecular-weight polymers.
- For the same molecular weight, linear polymers are generally better than branched or cross-linked polymers.
- In the case of polyelectrolytes there may be an optimum charge density.
- Ionic strength effects may be important, with a possible optimum value.
- Very strong flocs can be formed by polymer bridging, but floc breakage may be irreversible.

6.3.5 Charge neutralization and "electrostatic patch" effects

Particles in natural waters, effluents, and a wide range of industrial suspensions are negatively charged. In these cases the most effective flocculants are often found to be cationic polyelectrolytes, which adsorb strongly on negatively charged particles. An obvious question then is the role played by charge neutralization, rather than polymer bridging. There are some complicating factors, including kinetic aspects (see Section 6.3.6), but for fairly dilute suspensions it is often found that the optimum flocculant concentration is that needed to neutralize the particle charge. For instance, the flocculant concentration giving zero electrophoretic mobility (or zeta potential) is usually found to lie within the optimum dosage range.

If the role of the cationic polyelectrolyte is simply to neutralize particle charge, then there are important practical consequences. For instance, the charge density of the polyelectrolyte should be more important than molecular weight. The latter only has to be high enough to ensure complete adsorption. In practice, low-molecular-weight polyelectrolytes, such as polyDADMAC, are found to be effective. These adsorb strongly on negative particles and can neutralize their charge and hence cause destabilization. Adsorption is so strong in such cases that it continues beyond the point of charge neutralization and can give *charge reversal* and *restabilization,* as with specific counterion adsorption (see Chapter 4, Section 4.4.3). It follows that there will be an optimum dosage for the most effective flocculation. Also, the dosage range over which flocculation occurs depends on ionic strength.

Highly charged polyelectrolytes tend to adsorb in a flat configuration (i.e., with most of the segments in *trains;* see Figure 6.9). This reduces the likelihood of bridging interactions, so that charge neutralization may be the predominant mechanism of destabilization. For this reason, and based on terminology discussed in Section 6.1.1, additives such as polyDADMAC are often referred to as "coagulants" rather than "flocculants."

However, there are many cases where high-molecular-weight cationic polyelectrolytes are more effective, so it is likely that bridging interactions are important. This is especially the case where strong flocs are needed, as in sludge dewatering and some

mineral processing applications. The cationic charge of the polyelectrolyte is still necessary to give strong adsorption, but its long-chain nature is important to promote bridging interactions between particles. In these cases, the cationic charge density of the polyelectrolyte may not be so important, provided that strong adsorption occurs.

As well as simple charge neutralization and polymer bridging, there is another important mechanism that can be very important in practice. This is the so-called *electrostatic patch* effect.

When highly charged cationic polyelectrolytes adsorb on particles with a fairly low negative surface charge density, it is not physically possible for each surface charge to be individually neutralized by a cationic segment of the adsorbed chain. The average distance between charged surface sites may be significantly lower than the spacing between cationic sites on the polymer chain. So, although *overall* charge neutralization may occur, at a suitable polymer dosage there will be a *local* heterogeneity of charge, giving a *charge mosaic* or *electrostatic patch* arrangement, shown schematically in Figure 6.10.

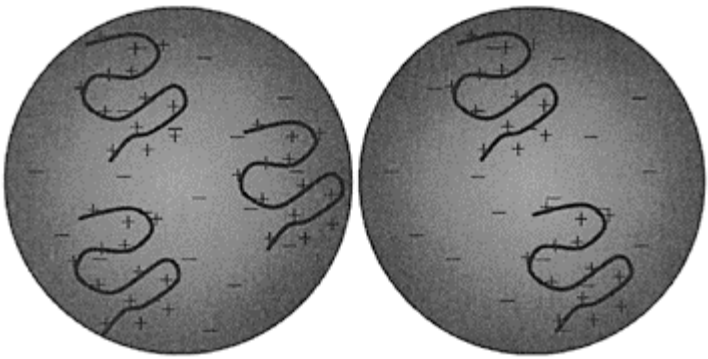

Figure 6.10 "Electrostatic patch" model for cationic polyelectrolyte and negatively charged particles.

The most important consequence of "patchwise" adsorption is that particles with no net charge can still show strong electrical attraction between regions of opposite charge. This extra attraction can give stronger flocs than those produced by simple salts or charge neutralization. These effects are dependent on ionic strength because increasing salt concentration causes attraction between oppositely charged patches to be reduced. However, there would also be a broadening of the dosage range within which flocculation occurs (see Figure 4.10).

6.3.6 Kinetic aspects of polymer flocculation

When a polymeric flocculant is dosed into a suspension of particles, several processes are initiated, the rates of which can greatly influence the flocculation behavior. Flocculants are usually added in the form of concentrated solutions (usually 0.1–1%) to give the

required final concentration, which is often of the order of mg/L. This means that the added polymer solution may be diluted by a factor of around 1000 during the dosing process. Because the polymer solution can be viscous, intense *mixing* is needed to achieve rapid and uniform distribution of the polymer molecules throughout the suspension. The polymer molecules also need to *adsorb* on particles before flocculation can occur. After adsorption, polymer chains undergo some rearrangement or *reconformation*, which would ultimately give an equilibrium conformation (as in Figure 6.9). Collisions of particles with adsorbed polymer occur, leading to the formation of aggregates or flocs. Finally, there is the possibility that flocs may undergo breakage under certain conditions. These processes are schematically illustrated in Figure 6.11 and are listed as follows:

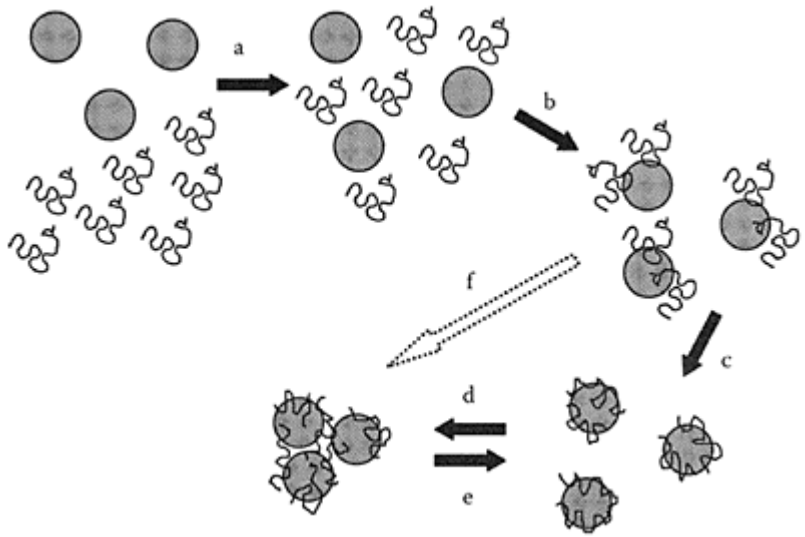

Figure 6.11 Schematic picture of different steps in polymer adsorption and flocculation (see text) tion (see text).

1. Mixing
2. Adsorption
3. Reconformation of adsorbed chains
4. Collisions to give flocs
5. Floc breakage

Although these are listed sequentially, it is likely that several occur simultaneously, which makes detailed analysis difficult. The rates of some of the individual processes can be estimated using simple models, and comparisons of these rates can be useful.

Mixing of polymer throughout the suspension is an important step and depends on the intensity of agitation during and after dosing. With inefficient mixing, it is likely that

local overdosing of polymer gives excess adsorption on some particles and insufficient adsorption on others. Excess adsorption may give restabilization of some particles, and this could be the reason why residual fine particles remain as a "haze" after flocculation and sedimentation. However, in the following we shall assume that mixing is sufficiently rapid to give essentially instantaneous distribution of added polymer throughout the suspension.

Adsorption of polymer can be treated as a collision process, the rate of which depends on transport phenomena. If the polymer molecule is assumed to be a random coil, of spherical shape, and with a certain hydrodynamic diameter and the particles are uniform spheres, then we can use the Smoluchowski approach to calculate collision rates (see Chapter 5). It is reasonable to suppose that every collision between a polymer molecule and a particle leads to adsorption, so that the collision rate gives the adsorption rate directly. Collisions of polymer molecules and particles are governed by diffusion and/or fluid motion, depending on the sizes of the colliding species and the shear rate.

Whatever the mechanism, the rate of collision between polymer molecules and suspended particles can be written in terms of a second-order rate process, as in Equation (5.1):

$$J_{12}=k_{12}N_1N_2 \tag{6.6}$$

where J_{12} is the number of collisions occurring in unit time and unit volume, k_{12} is a rate coefficient, and N_1 and N_2 are the number concentrations of particles and polymer molecules, respectively. (In most practical cases $N_2 >> N_1$.)

It is reasonable to assume that a certain fraction, *f*, of the added polymer needs to be adsorbed before the particles are adequately destabilized (either by bridging interactions or charge neutralization). The time required for this adsorption to occur can easily be derived from Equation (6.6), assuming that the particle concentration remains constant. (This assumption makes polymer adsorption a *first-order* rate process.) With increasing adsorption, the polymer concentration in solution is reduced, giving a progressive lowering of the adsorption rate. The characteristic time, t_A, for a fraction *f* of the added polymer to be adsorbed is given by the following:

$$t_A = -\frac{\ln(1-f)}{k_{12}N_1} \tag{6.7}$$

This expression shows that the adsorption time will be higher for low particle concentrations and independent of the initial polymer concentration. Also, it follows that, in principle, an infinite time would be needed for 100% of the polymer to be adsorbed (f=1), which is characteristic of first-order rate processes.

To calculate t_A, we need to have expressions for the rate coefficient, k_{12}. For transport by diffusion and fluid motion, or shear, these are Equations (5.6) and (5.20), respectively, which can be written as follows:

$$\text{Diffusion: } k_{12} = \frac{2k_BT}{3\mu}\frac{(d_1+d_2)^2}{d_1d_2} \tag{6.8}$$

$$\text{Shear:} \qquad k_{12} = \frac{G}{6}\left(d_1 + d_2\right)^3 \tag{6.9}$$

where d_1 and d_2 are the diameters of the particles and the polymer molecules, respectively

By assuming suitable values for the diameters and shear rate, G, we can calculate adsorption times from Equation (6.7) and the appropriate rate coefficient. Of course, adsorption by diffusion and fluid motion would occur simultaneously, but it is convenient to calculate the adsorption time for each mechanism as though the other were inoperative. We can then get some idea of their relative importance. The results in Figure 6.12 are for spherical particles of 1-μm diameter, with the effective polymer coil diameter varying from 10 to 1000 nm (corresponding to molecular weights in the range from a few hundred up to many millions). The particle concentration, N_1, is assumed to be $2\times10^{14}/m^3$, which is equivalent to a volume fraction of about 100 ppm for 1-μm particles. The results for the diffusion-controlled adsorption process are for aqueous systems at 25°C. Those for shear-induced adsorption are for shear rates of 50 and 500 s^{-1}, which are typical of values during flocculation and rapid mixing, respectively.

As expected, adsorption times for the diffusion-controlled process increase with increasing polymer size, whereas those for shear-induced adsorption decrease. Thus there is a polymer size at which the two processes are of equal significance. For G=50 s^{-1}, this is in the region of 300 nm, so except for quite large molecules, adsorption should mainly be controlled by

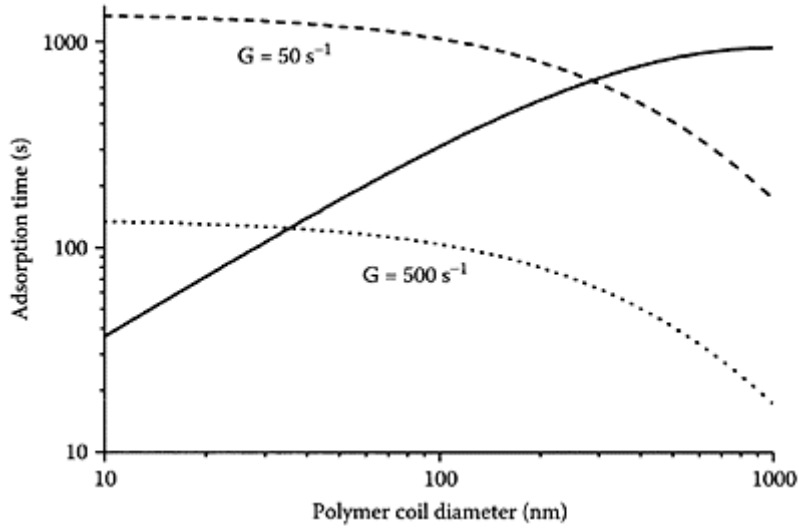

Figure 6.12 Calculated polymer adsorption times as function of polymer size for transport by diffusion *(full line)* and shear (*broken lines;* shear rates shown on curves). The

concentration of particles is $2\times10^{14}/m^3$,
and their diameter is 1 μm.

diffusion. However, for the higher shear rate, it appears that polymers larger than about 35 nm would be transported mainly by fluid motion. It is also worth noting that the actual adsorption times predicted are long—up to several minutes. Although the results are greatly dependent on the assumed parameters, these are reasonable for flocculation under typical water treatment conditions. Adsorption times would be lower at higher particle concentrations and, in the case of larger polymers, at higher shear rates.

Whether long adsorption times are practically significant depends on the collision rate between particles and the corresponding flocculation time.

Reconformation of adsorbed polymer chains probably occurs fairly rapidly, although not much is known about this aspect. For high-molecular-weight polymers, times of the order of a few seconds or longer may be needed to attain the final equilibrium. The reconformation rate should depend only on the nature of the polymer and the degree of surface coverage. It should be independent of the particle concentration, and this has important implications for nonequilibrium flocculation, which will be considered later in this section.

Flocculation of particles can be characterized by an average time between collisions. This follows directly from Equation (5.13), which can be rewritten as follows:

$$t_F = \frac{2}{k_{11}N_1} \tag{6.10}$$

Here, t_F is a characteristic *flocculation time*, which is equivalent to the coagulation time, τ, discussed in Chapter 5, Section 2.1. (The factor 2 is included because the collision rate coefficient, k_{11} is twice the aggregation rate coefficient, k_a.) The rate coefficient, k_{11}, is calculated by setting $d_1=d_2$ in Equations (6.8) and (6.9).

It is instructive to compare the calculated flocculation time with adsorption times, such as those in Figure 6.12. For the same particle concentration and shear rates, the following values for t_F and t_A are found (t_A values are for a polymer of 100-nm diameter):

Diffusion:	t_F=815 s	t_A=310 s
Shear, G=50 s^{-1}	t_F=150 s	t_A=1040 s
Shear, G=500 s^{-1}	t_F=15 s	t_A=104 s

For diffusion only (i.e., without applied shear) sufficient polymer would be adsorbed before most particles had experienced one collision, so that adsorption would not be a rate-limiting step. However, shear-induced particle collisions occur much more rapidly and the adsorption time is considerably longer than the flocculation time. For G=50 s^{-1} diffusion is the most effective polymer transport process, but the adsorption time (310 s) is about twice the flocculation time. For the higher shear rate, both adsorption and flocculation are dominated by shear, but t_A is about 7 times longer than t_F. This means that, on average, a particle would experience 7 collisions with other particles before sufficient polymer had adsorbed to give adequate destabilization. For this reason a significant lag time is often observed after adding a polymeric flocculant to a dilute

suspension before flocculation begins. For more concentrated suspensions than the one in our example, both adsorption and flocculation rates are higher and the lag time may not be of practical significance. However, conclusions concerning the *relative* rates and times are still valid.

Breakage of flocs may be important in agitated suspensions and has been discussed in Chapter 5, Section 5.4. In the case of polymer bridging there is some evidence that floc breakage at high shear rates may be irreversible to some extent, so that broken flocs do not readily reform when the shear rate is reduced. This is an important practical consideration in the application of polymeric flocculants.

Nonequilibrium flocculation is the process indicated by the broken arrow in Figure 6.11. It can be important when the reconformation of adsorbed polymer chains (c) is relatively slow compared to the particle collision rate. Because the latter depends on particle concentration, whereas reconformation does not, it follows that this effect will be more important for more concentrated suspensions. In such cases, particles may collide before their adsorbed polymer chains have attained their equilibrium conformation, when they are in a more extended state. This can lead to more effective bridging links between particles, so that, in more concentrated suspensions, stronger flocs may be formed. With cationic polyelectrolytes and negative particles the equilibrium adsorbed configuration is usually flat, so that bridging interactions are less likely. However, if particle collisions occur before significant "flattening" of adsorbed chains has taken place, then bridging may play a more important role than charge neutralization. It is likely that, in the flocculation of concentrated suspensions (of the order of 1% solids or more) with high-molecular-weight polymers, "nonequilibrium flocculation" plays a predominant role.

Although our discussion of kinetic aspects of flocculation has been entirely in terms of polymeric flocculants, it is likely that some of the concepts apply to other additives, especially hydrolyzing metal coagulants, where precipitated hydroxide particles play an important role.

6.3.7 Applications

Since their introduction in the 1950s, polymeric flocculants have been widely used in many industrial applications. Some of the main areas of application are as follows:

- Drinking water treatment
- Sludge conditioning and dewatering
- Mineral processing
- Papermaking
- Biotechnology

In all of these areas, synthetic polymers, often based on polyacrylamide, are much more widely used than natural materials. Depending on the application, different polymer characteristics may be required. The most important properties are ionic character and charge density, molecular weight, structure (linear or branched), and hydrophilic/hydrophobic nature. All of these characteristics can be controlled during synthesis, and, in principle, it would be possible to "tailor make" a polymeric flocculant to suit a specific application. In practice, a more empirical approach is usually adopted, in which a range of flocculants are tried and the most effective agents are chosen. It is rarely

possible to predict the most effective flocculant for a new application on the basis of fundamental principles.

In *water treatment* there are several potential applications of polymeric flocculants. They may be used as *primary flocculants* with no other additive. In this case, high-charge, low-molecular-weight cationic polyelectrolytes are most often used. Their role is to neutralize the charge of anionic impurities in water. Because the particle concentration is often low, flocculation can be slow and this may be a problem. Also, precise dosage control is needed to avoid overdosing and restabilization of particles. Cationic polyelectrolytes are also used in direct filtration applications because they can make capture of particles more effective (see Chapter 7).

Another application of polymers in drinking water treatment is in conjunction with hydrolyzing metal coagulants, in which case the additives are known as *coagulant aids.* The aim here is to strengthen metal hydroxide flocs, which are otherwise weak. For this purpose low-charge, often anionic poly-electrolytes of high molecular weight are found to be effective. The polymer is usually added shortly after the metal salt, when hydroxide precipitates are already formed.

For drinking water treatment, there are stringent requirements relating to the toxicity of additives. Relatively few products have been approved for this purpose.

In water and wastewater *sludge dewatering,* polymers can be effective in increasing both the dewatering rate and the solids content of the dewatered sludge. For this application high-molecular-weight cationic polyelectrolytes of moderate charge density are often found to be most effective, although the precise mechanisms of action are not well-understood. Some dewatering devices, such as high-speed centrifuges, generate very high shear rates, so strong flocs are needed to avoid breakage. Such processes only became practical when high-molecular-weight polymeric flocculants became available.

In *mineral processing* there are many solid-liquid separation operations where polymers play a vital part and a wide range of polymer types are used, depending on the specific application. Flocculants may be used to remove unwanted impurities, as in the flocculation of iron oxide particles in the Bayer alumina process. They are also widely used in dewatering operations. A novel application is *selective flocculation,* in which one component in a mixed suspension can be flocculated and separated. This process requires that the chosen polymer adsorbs on one type of particle but not on others, so that some specific adsorption interaction must be involved (see Section 6.3.3).

In *biotechnology* it is often necessary to remove microbial cells (usually bacteria or yeast) from a fermentation medium to process the cells to recover valuable products. Polymeric flocculants, mostly cationic polyelectrolytes, are extensively used in these applications. Some natural products, especially chitosan, may also be effective.

Papermaking applications of flocculants are mainly concerned with the retention of certain ingredients, such as filler particles, in the formed paper sheet. Polymers may also be useful in binding cellulose fibers in the sheet. Cationic polyelectrolytes, including cationic starch are used, and there are some examples where insoluble, highly cross-linked polymers are used as retention aids.

Further reading

Dentel, S.K., Coagulant control in water treatment, *Crit. Rev. Environ. Control*, 21, 41, 1991.
Duan, J. and Gregory, J., Coagulation by hydrolysing metal salts, *Adv. Colloid Interface Sci.*, 100–102, 475, 2003.
Gregory, J., Polymer adsorption and flocculation, in *Industrial Water Soluble Polymers*, Finch, C.A., Ed., Royal Society of Chemistry, 62, 1996.
Liss, S.N. (Ed.), *Flocculation in Natural and Engineered Systems*, CRC Press, Boca Raton, FL, 2004.
Richens, D.T., *The Chemistry of Aqua Ions*, Wiley, Chichester, 1997.

chapter seven
Separation methods

7.1 Introduction

In this chapter we shall briefly survey the main methods used to remove particles from water. Only a broad outline will be given, without going into a lot of technical detail. Most emphasis will be on processes used in drinking water treatment, although many of the concepts are relevant to other applications. The separation methods covered are as follows:

- Sedimentation
- Flotation (mainly dissolved air flotation)
- Filtration (including deep bed and membrane processes)

In all cases there is a strong influence of particle size, and it is often found that increasing particle size by a coagulation/flocculation procedure is a necessary preliminary step before one or more of the previously mentioned processes is used. Filtration can be an effective method, but, for various reasons, it may be preceded by another separation process, either sedimentation or flotation. This can greatly reduce the load on the subsequent filtration process and leads to longer filter runs. A typical sequence of steps in a solid-liquid separation procedure is shown schematically in Figure 7.1.

Although the principles of coagulation and flocculation have been dealt with in some detail in the previous chapter, some discussion of more practical aspects will be given here, followed by sections on the three processes listed earlier.

7.2 Flocculation processes

The main requirements for effective flocculation are as follows:

- Rapid mixing of coagulants
- Opportunity for collisions of destabilized particles and hence flocculation

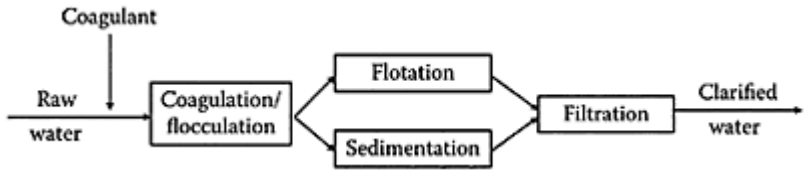

Figure 7.1 Typical sequence of processes for particle separation in a water treatment plant.

For the second step some form of fluid motion has to be generated, which may be by mechanical stirring or flow (or both).

7.2.1 Rapid mixing

Essentially rapid mixing (sometimes called "flash mixing") is necessary to distribute the coagulant species among the particles in as short a time as possible. In the case of coagulants that adsorb on particles and neutralize their charge, this can be especially important. Poor mixing can lead to local overdosing of coagulant and hence restabilization of some particles, as mentioned in Chapter 6, Section 6.3.6. For this reason, a short period of intense, turbulent mixing is desirable. The high shear rates associated with rapid mixing can also play an important part in the transport of coagulant species and can increase the rate of adsorption. In the case of hydrolyzing metal coagulants, under conditions where hydroxide precipitation and sweep flocculation are important, the role of rapid mixing is not so clear. However, it is known that hydrolysis rates are rapid and it is likely that rapid mixing conditions have some role in determining the relative rates of key processes such as adsorption and the formation of precipitates.

Ideally, rapid mixing needs to be intense but of short duration (no more than a few seconds). Otherwise, the nature of flocs formed subsequently can be affected. Prolonged periods of intense mixing can lead to the growth of small, compact flocs that grow slowly when the shear rate is reduced.

Rapid mixing may be carried out in a flow-through stirred tank (a "backmix" reactor), although this is an inefficient mixing device because of short-circuiting of flow. It is difficult to achieve complete and homogeneous distribution of added coagulant in a short time (say, less than 1 second). It is more common to add coagulant at a point where there are turbulent conditions as a result of flow. This point may be in a channel—for instance, where water flows over a weir—or in some kind of "in-pipe" mixer. The latter method can involve adding coagulant at a point where the pipe either widens or narrows, as shown schematically in Figure 7.2.

Although rapid mixing has long been recognized to have important effects on flocculation processes and has been studied in some detail, it is likely that many instances of poor performance of practical flocculation units can be attributed to inadequate mixing.

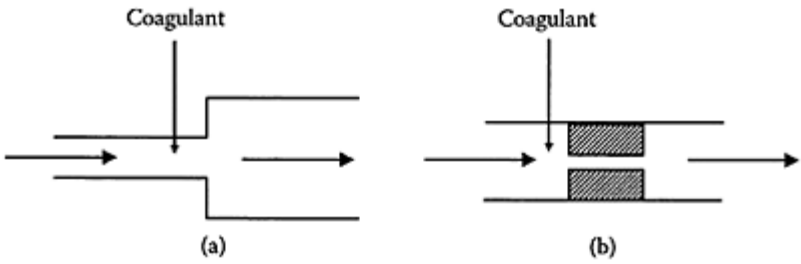

Figure 7.2 Rapid mixing of coagulant by "in-pipe" methods. (a) Widening pipe; (b) Narrowing pipe.

7.2.2 Floc formation

In most cases, growth of large flocs requires the application of velocity gradients or shear. The fundamental aspects of orthokinetic flocculation were considered in Chapter 5, Section 5.2.2. The major influences on flocculation rate are the particle (floc) size and concentration and the effective shear rate, G. Higher shear rates give enhanced particle collision rate but may reduce collision efficiency and cause some floc breakage. A useful compromise is a process known as *taper flocculation*, in which the effective shear rate is initially high, giving a rapid flocculation rate, and then progressively reduced so that large flocs can form.

In practice, application of shear involves the input of energy. This can be achieved in essentially two ways: *mechanical* or *hydraulic*.

Mechanical devices are typified by flow-through stirred tanks of various kinds, sometimes known as *paddle flocculators*. The paddles may rotate about vertical or horizontal axes, but in all cases the power input to the water depends on the drag force on the paddle and the rotation speed. The power input to the water could, in principle, be measured, but it is not too difficult to calculate. The power transferred from a moving paddle to water is simply the drag force multiplied by the paddle velocity (relative to the water). The drag force (see Chapter 2, Section 2.3.1) is given by the following:

$$F_D = \frac{1}{2} C_D \rho_L \left(v_p - v\right)^2 A_p \qquad (7.1)$$

where $(v_p - v)$ is the relative velocity of the paddle blade to the water and A_p is the projected area of the blade normal to the motion. The drag coefficient C_D depends on the shape of the paddle blade, but it is usually in the range of 1–2.

The power input to the water is as follows:

$$P = \frac{1}{2} C_D \rho_L \left(v_p - v\right)^3 A_p \qquad (7.2)$$

It is then possible to calculate the power input per unit mass of water, ε, and hence to calculate an effective shear rate using Equation (5.26).

Alternatively, if the power input to the motor driving the paddle is known, as well as the efficiency (the fraction of power actually transmitted to the water), then we can calculate the energy dissipation directly. For a water volume of 400 m^3 and a motor with a power of 1 kW and an efficiency of 60%, the effective shear rate turns out to be about 40 s^{-1}.

Flow-through flocculation tanks may contain several paddles in sequence, and taper flocculation can be achieved by arranging for the rotation speed of successive paddles to be progressively reduced. Average shear rates are usually in the region of 20–70 s^{-1}, and residence times in the tank may be of the order of 20 minutes. For this residence time and an average shear rate of 50 s^{-1}, the Camp number, *Gt*, is 60,000, which is characteristic of simple flow-through flocculators.

Hydraulic flocculators rely on flow to provide velocity gradients. Because of fluid drag, there is an inevitable dissipation of energy, which is manifested as a pressure difference or *head loss, h*. If the volume flow rate through the flocculator is Q, then the power dissipated is as follows:

$$P=\rho_L g Q h \tag{7.3}$$

where *g* is the acceleration as a result of gravity.

Hydraulic flocculation occurs as a result of flow in pipes. At very low flow rates, or in narrow tubes, laminar conditions apply and it can be shown that the Camp number, *Gt*, takes a simple form:

$$Gt = \frac{16L}{3D} \tag{7.4}$$

where *L* is the length and *D* is the diameter of the tube.

It is noteworthy that the *Gt* value depends only on the dimensions of the tube and not on the flow rate. This is because the average shear rate increases linearly with flow rate, whereas the residence time in the tube is inversely proportional to flow rate. Thus, the flow rate has no net effect on *Gt*.

Whereas laminar tube flow can be useful in laboratory flocculation tests, practical tube flocculators always operate under turbulent conditions (for Reynolds numbers greater than about 2000), where Equation (7.4) does not apply. For turbulent pipe flow the head loss is given by the *Darcy-Weisbach* equation:

$$h = \frac{2fLv^2}{gD} \tag{75}$$

where *v* is the average velocity in the pipe (=4Q/π*D*2) and *f* is the *friction factor*, which depends on the Reynolds number and the roughness of the

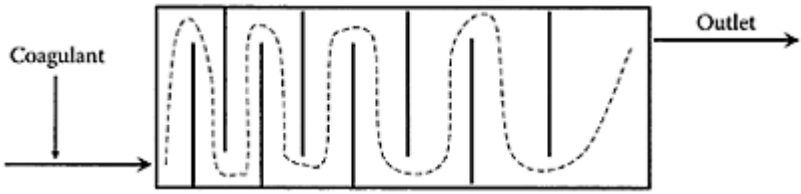

Figure 7.3 Schematic diagram of a baffled tank flocculator. Note that the baffles become more widely spaced toward the outlet, giving lower effective shear rates and *taper flocculation*.

pipe. The friction factor, for various conditions, is presented graphically in many textbooks on fluid mechanics.

It turns out that G values in the required range for flocculation can easily be achieved by turbulent flow in pipes. The problem is that residence times of the order of 20 minutes are needed, and, for reasonable flow rates, this corresponds to very long pipes (typically of the order of 500 m). For this reason pipe flocculators are not generally practical in water treatment, although existing pipes may be useful in providing some flocculation.

A better alternative is some form of *baffle flocculator*, which consists of a channel or tank with an arrangement of baffles, so that the flow undergoes several changes of direction (Figure 7.3). This can give significant head loss and hence appreciable G values, whereas sufficient residence time can be achieved in tanks of manageable size. Taper flocculation can be achieved by changes in the shape or spacing of successive baffles.

Hydraulic flocculation may also occur in flow-through packed beds, as *in deep bed filtration*, or in fluidized beds, as in *upflow clarifiers*. These will be dealt with briefly in the following sections.

7.3 Sedimentation

7.3.1 Basics

Fundamental concepts of sedimentation were covered in Chapter 2, Section 2.3.3. For a dilute suspension of small particles, *Stokes Law*, Equation (2.29) is applicable, so that settling rate depends on the square of the particle size and the effective (buoyant) density. However, in what follows we do not need to restrict discussion to Stokesian particles. For larger particles, the settling rate is determined by particle size and density and there is a characteristic *terminal velocity*, which is rapidly established. Because there is generally a distribution of particle size, there will be a corresponding distribution of settling velocities. This can be determined by an experimental *batch settling test*, which may give results like those in Figure 7.4. This shows the proportion of particles, *f*, with a settling

velocity smaller than a given value, v. (The significance of the terms f_0 and v_0 will be explained in the next section.)

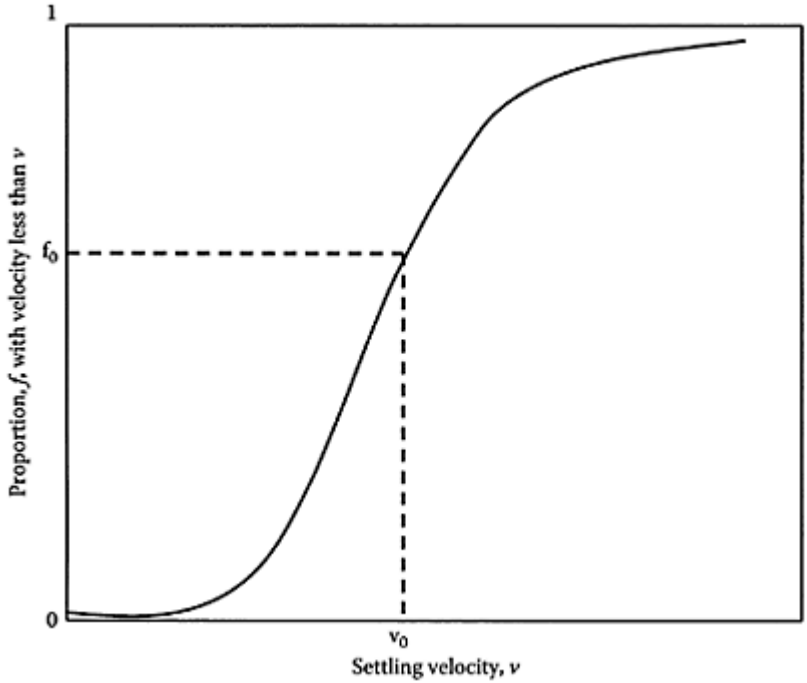

Figure 7.4 Distribution of settling velocities from a *batch settling test*.

7.3.2 Sedimentation in practice

Practical sedimentation units take many forms. The simplest is a *batch tank*, which has to be filled and emptied for each operation. It is much more convenient to use a flow-through vessel, and it is easiest to consider a rectangular tank with horizontal flow, which may be regarded as an *ideal settling basin* (Figure 7.5). It is assumed that suspension enters the tank with a uniform concentration throughout the inlet zone and that flow occurs uniformly in a horizontal direction. This is the so-called *plug flow* condition, where all elements of fluid have the same velocity and hence the same residence time in the tank. At the bottom of the tank is a sludge zone, and it is assumed that all particles reaching this zone are permanently removed from the suspension. All particles that do not reach the sludge zone during their passage through the tank are assumed to leave at the outlet zone.

There is certain critical settling velocity, v_0, such that all particles settling faster than this value will be removed. This is easily calculated from the height of the settling zone, H, and the residence time, τ. The latter depends on the volumetric flow rate, Q, and the volume of the settling zone, HA, where A is the surface area. Particles with a settling

velocity, v_0, entering at the top of the inlet zone will *just* reach the sludge zone, as shown in Figure 7.5. Thus:

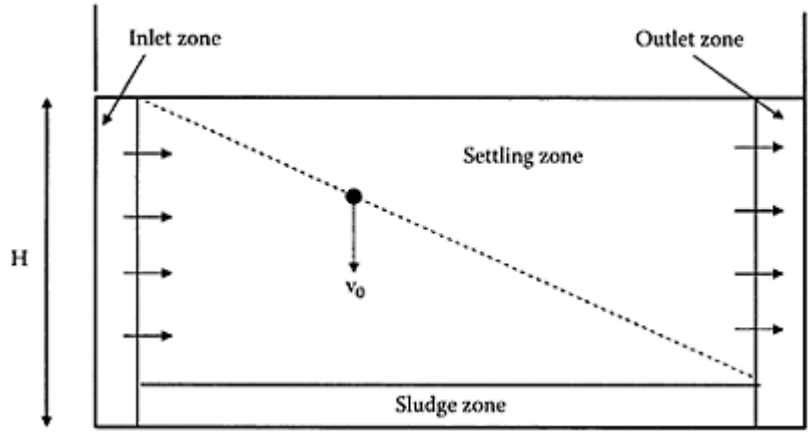

Figure 7.5 An ideal settling basin.

$$v_0 = \frac{H}{\tau} = \frac{HQ}{V} = \frac{Q}{A} \tag{7.6}$$

The term Q/A is known as the *surface loading rate* or *overflow rate* and is equal to the critical settling velocity. All particles with this settling velocity, entering at the top of the inlet zone, will just be removed during its passage through the settling zone. Particles with a smaller settling velocity may also be removed if they enter at a lower position (see Figure 7.5). Note that the critical settling rate for a given flow rate depends on the surface area of the tank and not the depth. Clearly, the larger the surface area, the lower the v_0 and hence a greater proportion of particles will be removed. (Of course, for a given volume flow rate, increasing surface area implies a decreasing depth.)

For particles with a settling velocity, v_s $(<v_0)$, a fraction of them, v_s/v_0, will be removed from the settling zone. A fraction $1-f_0$ of particles have settling rates greater than or equal to v_0, and all of these will be removed. So, the total fraction of particles removed is given by the following:

$$F = (1 - f_0) + \int_0^{f_0} \left(\frac{v_s}{v_0} \right) df \tag{7.7}$$

Although this expression gives a useful guide to the behavior of settling tanks, the assumptions made, such as plug flow and uniform inlet concentration, mean that quantitative predictions will be subject to some uncertainty in practical applications.

In addition to rectangular sedimentation tanks, radial flow designs are also common and have some hydraulic advantages. However, conventional plant-scale sedimentation requires tanks of quite large area because of the

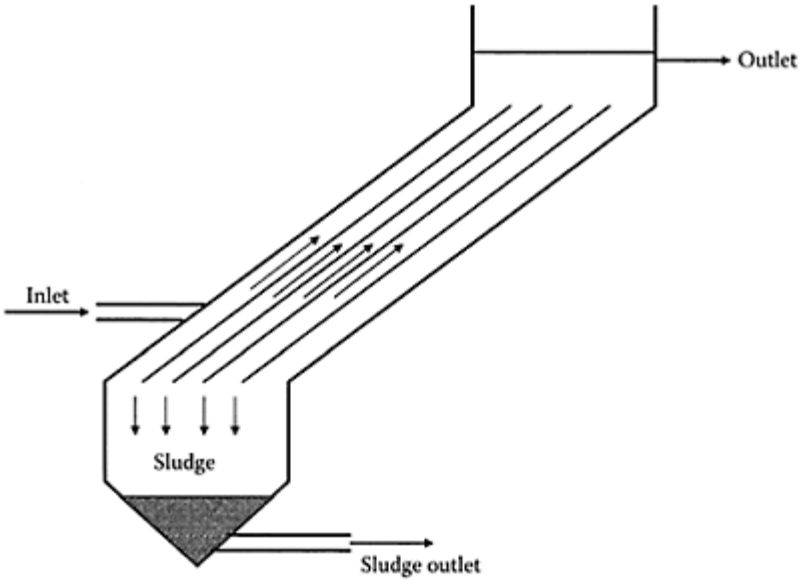

Figure 7.6 An inclined plate separator.

need to maintain the appropriate surface loading rate (typically of the order of 1–2 m/h).

There are ways to reduce the required area, including the use of stacked horizontal trays, but it is more convenient to use an *inclined plate separator*, shown schematically in Figure 7.6 These provide more surface for sedimentation in a given plan area (effectively several shallow settling basins in parallel). Particles settling on the plates accumulate as sludge, which slides by gravity to a collection zone. *Tube settlers* operate on a similar principle.

7.3.3 Upflow clarifiers

By flowing a coagulated suspension upward through suitable tank, it is possible to achieve a condition where flocs settle at a rate equal to the upflow velocity of the water, thus creating a *floc* or *sludge blanket* (Figure 7.7). Effectively, incoming destabilized particles pass through a fluidized bed of preformed flocs; this gives a greatly enhanced flocculation rate. According to Equation (5.24), the rate of orthokinetic flocculation is directly proportional to the solids concentration, and this is much higher in the floc blanket than in the incoming water.

Another point is that floc growth in the blanket is by the attachment of small particles to existing flocs, which gives denser flocs than those produced by cluster-cluster aggregation (see Chapter 5, Section 5.3.1). This means that the flocs will have a higher settling rate, so higher upflow rates are possible.

The combination of flocculation and sedimentation in a single clarifier unit has great advantages. There are many different commercial designs of flocculator-clarifiers, and these are widely used in practice.

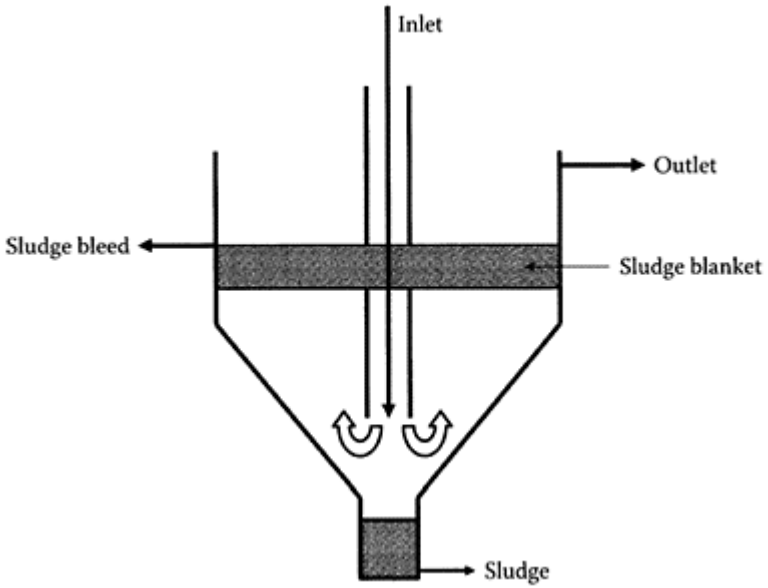

Figure 7.7 Schematic diagram of an upflow clarifier.

7.4 Flotation

7.4.1 General

Flotation is a process whereby particles become attached to air bubbles that rise to the surface, thus removing particles from suspension. This process is of enormous practical and economic importance, especially in the mineral industry, where billions of tons of ore are treated annually by flotation.

For an air bubble to attach to a particle in water, the particle must be *hydrophobic* (water-repelling) to some extent and hence have a finite *contact angle* with water. Water spreads completely on a hydrophilic surface, but it forms a contact angle if the surface has some hydrophobic character. Some minerals are hydrophobic and *naturally floatable*. These include many sulfide minerals, talc, and graphite. However, most minerals are hydrophilic and can only be floated if their surface is modified by certain reagents, generally known as *flotation collectors*.

In mineral processing the primary use of flotation is to separate minerals from mixtures (i.e., *selective flotation*). This exploits the different floatability of different components of the mixture. Usually, the ore is ground, with water and appropriate reagents, down to some chosen grain size. The finest particles or “slimes” (less than about 20 μm in size) are separated out for treatment and the coarser particles are treated by flotation with air bubbles. Air is usually introduced by a stirrer, which also generates bubbles. With the right choice of reagents and other chemical conditions it is possible to

make some components of the mixture easily floatable and others much less so. The floated particles rise as a froth (the process is often called *froth flotation*) and can be removed by skimming, usually followed by further purification stages. The froth flotation process is commonly used around the world, especially in the production of metals, and makes possible the use of low-grade ores, which would otherwise be difficult to treat.

When air bubbles are introduced by a mechanical process, as in froth flotation, the process is called *dispersed air flotation*. The bubbles produced are large (up to a few millimeters), but these are appropriate for removing the coarse and dense particles encountered in mineral processing. There are other methods available that produce finer bubbles, such as *electrolytic flotation* and *dissolved air flotation*, which are better suited to water and waste-water treatment.

Electrolytic flotation or *electro-flotation* involves passing a direct current between suitable electrodes in water to generate hydrogen and oxygen bubbles. Although attractive in principle, this process is uneconomic and has a number of disadvantages. For water treatment, dissolved air flotation (DAF) is much more widely used and will be discussed in the next section.

7.4.2 Dissolved air flotation

Dissolved air flotation is a fairly common process in water and wastewater treatment. In water treatment it is especially useful for removal of particles with low density, such as algae, which can be difficult to separate by traditional sedimentation methods even after flocculation (because of the fractal nature of flocs; see Chapter 5, Section 5.3.3).

The most common mode of operation is to saturate part of the water with dissolved air at high pressure. The saturated water is injected into the main water flow, containing preformed flocs, and the sudden reduction of pressure causes air to be released as fine bubbles. The bubbles attach to flocs, which then rise to the surface as a float layer, leaving clarified water below. The water that is saturated with air is usually taken from the clarified stream, giving *recycle-flow* DAF. A schematic diagram of a DAF plant is shown in Figure 7.8.

Air has only limited solubility in water—about 25 mg/L at atmospheric pressure (1 bar) and 20°C. However, under practical conditions, the solubility is governed by *Henry's law*, which states that the solubility of a gas in a liquid is linearly proportional to the gas pressure. Thus, by increasing the air pressure, the solubility can be increased. At a typical operating pressure of 5 bar, the solubility of air in water would be 6 times that at atmospheric pressure (because the applied pressure is in excess of atmospheric). In DAF plants air is dissolved in water at high pressure in a *saturator*, often a packed column to give efficient contact between gas and liquid. In practice, around 90% saturation can be achieved (i.e., about 90% of the theoretical solubility predicted by Henry's law).

The pressurized water is introduced into the *contact zone* (see Figure 7.8) through a valve or nozzle, giving a sudden reduction of pressure and an immediate release of fine air bubbles, usually in the size range of 30–100 μm. It is generally found that the higher the pressure, the smaller the bubbles,

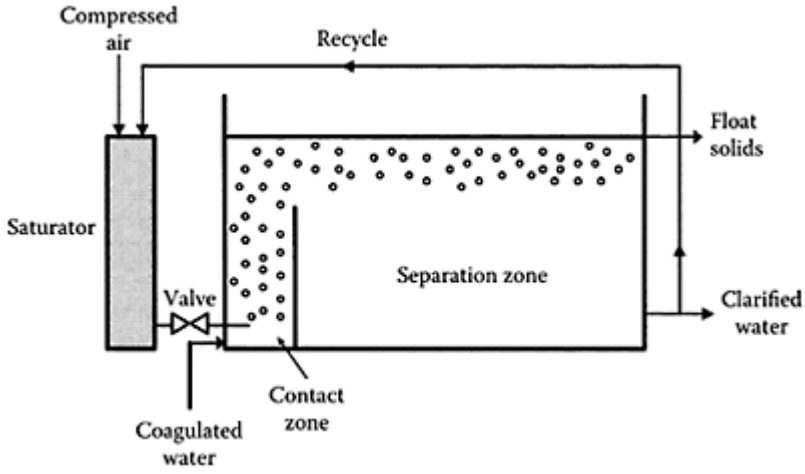

Figure 7.8 Schematic diagram of a dissolved air flotation (DAF) treatment process.

although the effect is not great. The concentration of air released in the contact zone depends on the amount of air dissolved in the saturator (and hence on the applied pressure) and the *recycle ratio*, which is usually in the range of 6–10%. If we assume saturation at 5 bar pressure, with 90% efficiency at 20°C, then the concentration of air in the water injected is 25×6×0.9 = 135 mg/L. So the amount of air released when the pressure is reduced is 135−25=110 mg per L of injected water from the saturator. Because this is diluted by a factor of 92/8, the concentration of air bubbles in the contact zone will be about 9.6 mg/L, which is in the required range for many water treatment applications. From the known density of air at 20°C (about 1.2 g/ L), we can calculate the volume concentration of air in water as about 8 mL/ L or a volume fraction of 8000 ppm. For bubbles of average diameter 50 μm, this corresponds to a number concentration of 1.2×10^8/L.

For a bubble to attach to a particle (or floc) it is essential for bubble-particle *collisions* to occur. A collision may or may not be effective in leading to attachment, depending on the interactions between particle and bubble. The particle surf ace needs to have some hydrophobic character, otherwise bubble attachment cannot occur. Attachment may also be hindered if both bubble and particle surfaces carry the same sign of charge, because of electrical repulsion (see Chapter 4, Section 4.3.2). Bubbles and most natural particles in water have negatively charged surfaces, so attachment may be highly improbable. By use of suitable coagulants, such as hydrolyzing metal salts, the surface charge of particles can be reduced and, in the case of hydrophilic particles, their surfaces may also be rendered more hydrophobic. Both of these effects will make bubble attachment more likely. (Also, the larger size of flocs makes collisions with bubbles more frequent; see later.) The interactions between particles and bubbles determine the value of the *collision efficiency*, which is the proportion of bubble-particle collisions resulting in attachment. This is analogous to the concept of collision efficiency in colloid stability (see Chapter 4, Section

4.4.4). With proper choice of coagulant and dosage, the collision efficiency should be not much less than 1.

The collision frequency between particles in suspension and rising bubbles depends on several factors, especially bubble and particle size and concentration. The mechanisms of particle capture by bubbles are similar to those that are important in deep bed filtration (see Section 7.5.1)—that is, *diffusion*, *interception*, and *sedimentation*. These will be considered more fully in the next section, and only qualitative conclusions will be given here:

- For bubbles of around 50 μm size, there is a minimum in capture efficiency for particle sizes in the region of 1 μm. Larger and smaller particles are captured more efficiently. This is one reason why flocculation is important for particles such as algae, with diameters of a few micrometers.
- Other things being equal, smaller bubbles give greater capture efficiency
- The higher the bubble concentration (or the recycle ratio) the better the removal of particles. However, there are practical limitations that restrict the recycle ratio to no more than about 10% in most cases.

For a given suspension of particles, there is a critical amount of air necessary to *just* prevent the particles from settling. This is easily calculated by considering the gravitational force on a particle in water, given by Equation (2.28). For a bubble, the corresponding force is as follows:

$$F_g = \frac{\pi d_b^3}{6}(\rho_b - \rho_w) \tag{7.8}$$

where d_b is the diameter of the bubble and ρ_b and ρ_w are the densities of the bubble and water, respectively. In fact, because the density of air is much smaller than that of water, ρ_b can be neglected in this expression without too much error.

By equating the upward force on the bubble with the downward force on the particle (assuming that the particle is denser than water), we can calculate the critical bubble size, d_{bc}, relative to the particle diameter, and the corresponding volumes. This gives the following:

$$\left(\frac{d_{bc}}{d_p}\right)^3 = \frac{V_{bc}}{V_p} = \frac{\rho_p}{\rho_w} - 1$$

where d_p and ρ_p are the diameter and density of the particle and V_{bc} and V_p are the volumes of the bubble and particle.

Finally, the critical air/solid ratio in terms of *mass* can be calculated. This is as follows:

$$\frac{\text{Air}}{\text{Solid}} = \left(\frac{\rho_p - \rho_w}{\rho_w}\right)\frac{\rho_b}{\rho_p} \tag{7.10}$$

It is worth noting that the reasoning leading to Equation (7.10) makes no assumptions about the size or shape of particles; we are only concerned with their buoyant mass. The

critical mass of air needed to prevent sedimentation is the same, regardless of the state of aggregation of the particles. In principle, the minimum concentration of air required to float particles should be directly proportional to the solids concentration. However, for particles of the order of a few micrometers in size the critical bubble size turns out to be much smaller than can be achieved in practical DAF plants. To float small particles, at least one bubble has to attach, and this represents a much higher air/solids ratio than given by Equation (7.10), especially when the particles have low density. With small particles and larger bubbles, it is possible that several particles will attach to one bubble, rather than just one, as assumed here, thus reducing the required air/solid ratio. However, in effect, a large excess of air would still be needed for small particles, which is an important reason why flocculation is needed ahead of DAF. This is illustrated in Figure 7.9, where the critical air/solids ratio is plotted as a function of floc diameter. For these calculations we need to assume values for the primary particle size and density, the fractal dimension of the flocs (see Chapter 5, Section 5.3) and an average bubble diameter. The values chosen are d_p=2 μm, ρ_p=1.1 and 2.5 g/cm^3, d_F=2.2 and d_b=50 μm. The density values are appropriate for biological particles, such as algae, and denser particles, such as clays.

The results in Figure 7.9 show a sharp decline in the critical air/solid ratio as the floc size increases and a constant value above a certain floc size (indicated by arrows on the figure). The constant value is the theoretical result predicted by Equation (7.10). For the denser particles this value is 7.2×10^{-4} and the floc size is 134 μm. At this floc size, the theoretical air/solid ratio corresponds to exactly one 50-μm bubble per floc. For the lower density particles, the corresponding values are 1.09×10^{-4} and 460 μm. (Allowing for more than one bubble per floc would change the nature of these results for small flocs, but the constant value for large flocs would be the same.)

In practice, higher bubble concentrations than the critical values from Equation (7.10) are needed, so that the bubble-floc aggregates rise at an appreciable rate. With the achievable bubble concentrations quoted earlier (around 10 mg/L) and particle concentrations up to 100 mg/L, the air/solid ratio would be at least 0.1, which should give a large excess of air. This gives rapid bubble-particle collisions and hence effective removal of particles.

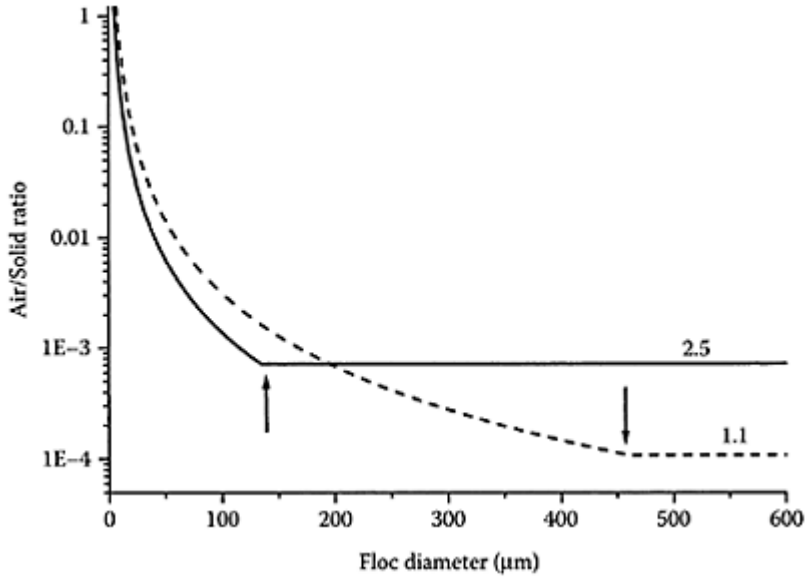

Figure 7.9 Critical air/solid ratio to just float flocs, as a function of floc diameter. The assumed densities of the primary particles are 2.5 and 1.1 g/cm^3, as shown on the curves. For other conditions, see text.

In the separation zone of a DAF plant (Figure 7.8) high rise rates can be achieved (in excess of 10 m/h), which is considerably higher than typical sedimentation rates of flocs. This means that the hydraulic loading rate can be significantly higher than that for a conventional sedimentation unit, giving reduced space requirements. DAF plants can operate at loading rates up to around 15 m/h, which makes the process attractive.

7.5 Filtration

There are two broad classes of filtration process used in solid-liquid separation. One involves flowing a suspension through a bed of granular material, such as sand, and is known variously as *deep bed filtration*, *depth filtration*, or *granular media filtration*. The other is *membrane filtration*, which is essentially a straining process by a thin layer of material with pores of a certain size.

7.5.1 Deep bed filtration

Deep bed filtration by far the most common filtration process used in water treatment. It has been used routinely since the 19th century and was a major advance in public health because filtration provides an effective barrier against many pathogenic microbes. Early filters operated at low *approach velocities* (up to about 0.5 m/h) through beds of fine sand. These have become known as *slow sand filters* and can be very effective. Their effect is a result of the fineness of the sand grains and biological action in the surface layer. Aerobic bacteria produce extracellular polymers, forming an adhesive network (the *Schmutzdecke*), which gives enhanced removal of fine particles. Because of the large land area required for slow sand filters and an inconvenient cleaning method, they are not widely used (London is a significant exception). We shall focus attention here on *rapid filters*.

Rapid filtration through granular media *(rapid sand filtration)* operates at approach velocities in the range of 5–30 m/h and relies principally on physical removal mechanisms. Flow may be by gravity *(rapid gravity filters)* or by applied pressure *(pressure filters)*. However, the same basic principles apply to both modes of operation.

A fundamental point is that rapid filtration uses granular media with typical grain sizes in the range of 0.5–2 mm. The pores between the grains are of the same order of size, which is usually much larger than the particles to be removed. This means that simple straining is not a significant removal mechanism. Rather, particles are removed by *deposition* on filter grain surfaces or on existing deposits of particles. There are two essential requirements for particle removal in deep bed filters:

- Particle *transport* to filter grains
- *Attachment* of particles to grain surfaces or to existing deposits

These are analogous to the collision and attachment steps in particle aggregation (Chapter 6), and similar principles apply. Transport depends on a number of f actors, including flow rate, grain size, particle size, and porosity of the bed. Attachment depends on the interactions between particles and grain surfaces, which are essentially the same colloid interactions that were discussed in Chapter 4. It is again appropriate to think in terms of a *collision efficiency*, α, which is the fraction of particles colliding with a grain surface that actually stick. With the right chemical conditions, repulsion between particles and filter grains can be eliminated, so that every collision results in attachment ($\alpha=1$). For simplicity, we shall make this assumption here.

In theoretical treatments of particle capture in deep bed filters it is common to refer to filter grains as *collectors*, and it is much easier to deal with spherical particles and spherical collectors. In practice, flow in filter pores is always slow enough for laminar conditions *(streamline flow)* to apply, and this makes the treatment easier.

An important concept for particle capture by collectors is the *single collector efficiency*, η. This is defined as the fraction of particles approaching the collector that collide with it. For a spherical collector the number of particles approaching the collector is given by the approach velocity of water, U (i.e., the velocity well upstream of the collector), the number concentration of particles, and the projected area of the collector. Because streamlines are diverted around the collector (Figure 7.10), the collector

efficiency can be much less than 1. This is a hydrodynamic effect, whereas the collision efficiency, α, is determined by colloid interactions and hence the chemistry of the system. Because we are assuming that α=1, the collector efficiency is

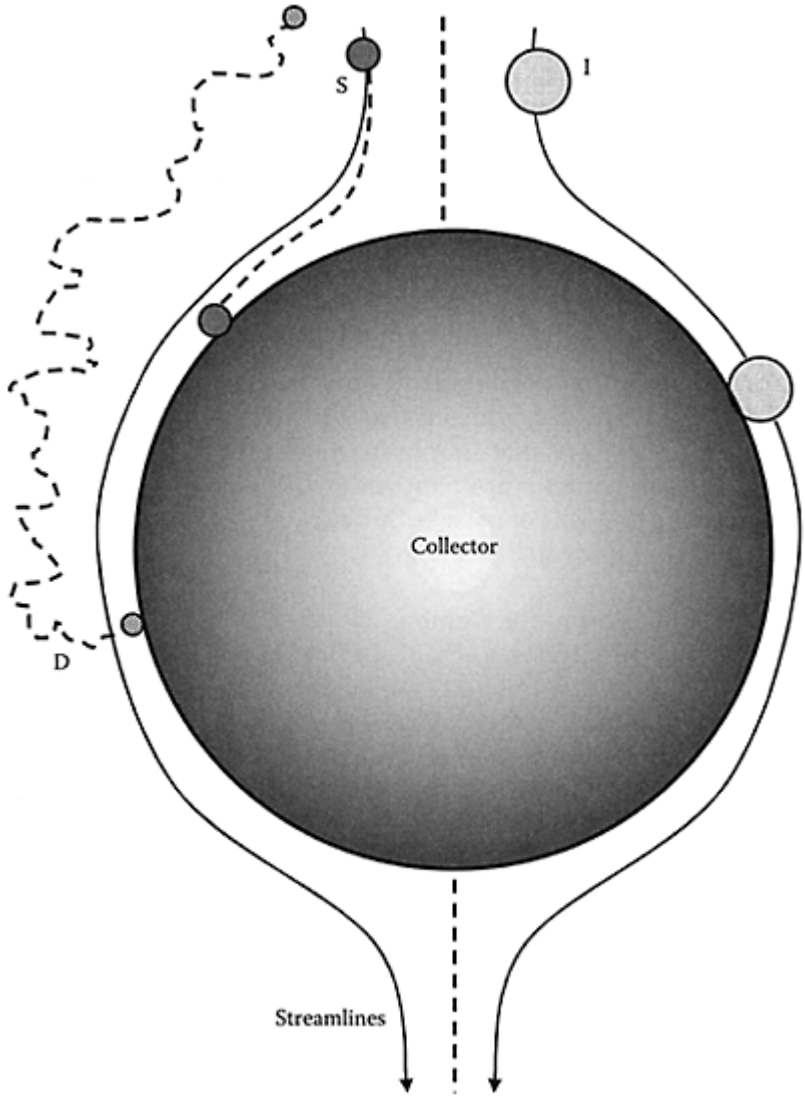

Figure 7.10 Capture mechanisms in deep bed filtration. Particles may contact the collector as a result of diffusion (D), interception (I), or gravitational settling (G). Diffusion

and settling cause the particles to depart from fluid streamlines.

the same as the *capture efficiency*. For spherical collectors, the collector efficiency is given by the following:

$$\eta = \frac{4I}{\pi d_C^2 Uc} \tag{7.11}$$

where I is the number of particles captured in unit time, d_C is the collector diameter, and c is the particle concentration.

There are three important particle transport mechanisms in water filtration, all of which are illustrated schematically in Figure 7.10:

1. *Diffusion*, where particles deviate from fluid streamlines as a result of brownian motion.
2. *Interception*, in which a particle is brought into contact with the collector if its center is on a streamline that passes within one particle radius of the collector surface.
3. *Sedimentation*, where gravity is responsible for the departure of particles from fluid streamlines to the collector surface.

These depend on the sizes of the particles and collector and on the fluid flow rate. The dependence on particle size is especially important because this affects the different transport mechanisms in different ways. Diffusion becomes more important for smaller particles, whereas the opposite is true for the interception and sedimentation. Although it is possible to derive analytical expressions for collector efficiency in all three cases, these involve simplifying assumptions and do not properly take into account important hydrodynamic interactions between particle and collector. (Similar effects are important in particle aggregation; see Chapter 5, Section 5.2.5.) A better approach is to carry out numeric computations to give collector efficiencies for the three transport modes, η_D, η_I, and η_G. These computations include the influence of van der Waals attraction between particles and collector, so that a value of Hamaker constant is needed (see Chapter 4, Section 4.2.3). Although the computations are for a single collector, the application is to grains in a packed bed filter. Allowance for the effect of neighboring grains can be made by including a coefficient that depends on the *porosity* of the packed bed (see later). Collector efficiency is increased by the presence of neighboring collectors, essentially because streamlines become more "crowded" and hence pass closer to the collector.

It is reasonable to suppose that the three contributions are additive, so that the total collector efficiency is given by the following:

$$\eta_T = \eta_D + \eta_I + \eta_G \tag{7.12}$$

Results of such computations, plotted against particle size, are shown in Figure 7.11, for the following conditions: grain diameter 0.5 mm, fluid velocity 5 m/h, particle density 2000 kg/m^3, temperature 20°C, Hamaker constant 10^{-20} J, and bed porosity 0.40.

The most striking feature of these results is the minimum in total collector efficiency for a particle size of around 1 μm. This is a consequence of the different effects of

particle size on the individual collector efficiencies. For very small particles, the dominant transport mechanism is diffusion, so the efficiency increases as the particle size decreases. For larger particles, interception and sedimentation are more significant and increasing particle size

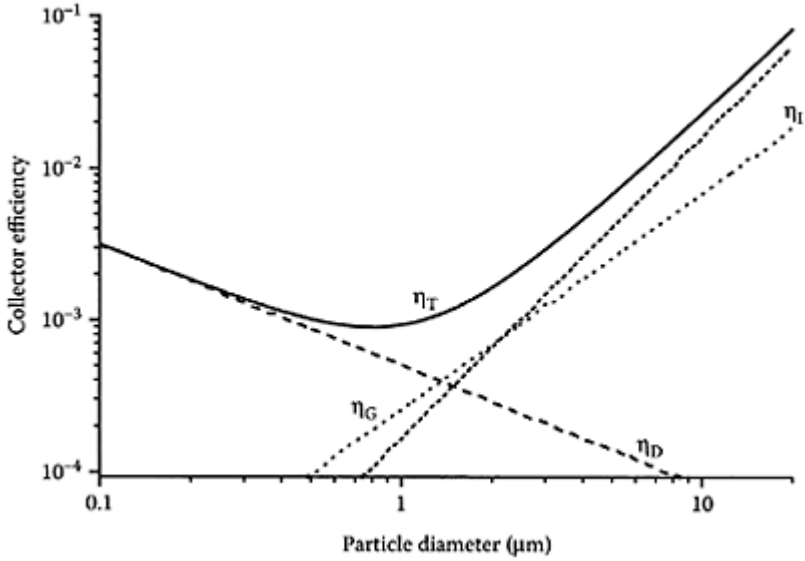

Figure 7.11 Computed collector efficiencies for a spherical collector, by diffusion, interception, and sedimentation, as a function of particle size. The total collector efficiency, ηT, is the sum of the individual contributions. For conditions, see text.

causes an increase in collector efficiency. For particle size in the region of 1 μm, diffusion is of about the same significance as the other transport mechanisms, and this is where the minimum in total collector efficiency occurs.

Flow rate is also an important parameter, as shown in Figure 7.12, where the total collector efficiency is shown for the same conditions as in Figure 7.11, except that the values are plotted for three different approach velocities. As this is increased from 1 to 20 m/h, the total collector efficiency decreases by a factor of around 10 for smaller particles, although the effect is less for larger particles. Increasing velocity causes particles to be swept past a collector more rapidly, giving less chance of capture. The higher shear rate also decreases the chance of capture. These effects are of great practical significance for deep bed filtration.

As a suspension flows through a packed bed, particles are removed by the mechanisms discussed earlier. so the particle concentration decreases with depth. This is often

described as a first-order process (removal proportional to particle concentration), with a characteristic *filter coefficient*, λ. This is defined in terms of the rate of change of concentration with depth, L:

$$\frac{\partial c}{\partial L} = -\lambda c \qquad (7.13)$$

(The right-hand side is written in partial differential form because time is another important variable.)

To proceed further we need to assume that the filter bed is of uniform composition, with no change of grain size with depth. In the early stages of a filter run, when the grains are still relatively clean, the filter coefficient can then be regarded as constant throughout the bed. In this case, Equation (7.13) can be integrated to give the concentration of particles emerging from a filter of depth *L:*

$$c = c_0 \exp(-\lambda_0 L) \qquad (7.14)$$

where c_0 is the concentration of particles in the suspension entering the filter (at depth L=0) and λ_0 is the *clean bed filter coefficient*. For a filter coefficient of 1 m^{-1}, the particle concentration decreases by a factor 1/*e* (about 0.37) for each meter of depth. Thus, for a 1-m deep filter the concentration of particles in the filtrate would be about 37% of the inlet concentration, or about 63% removal. For a filter depth of 2 m, the corresponding removal would be about 86%. Note that, in principle, 100% removal is not attainable in a filter of finite depth. In other words, we are not dealing with an *absolute* filtration method, where all particles above a certain size are removed. However, it is feasible to approach 100% removal quite closely in some cases.

There is a simple relationship between λ_0 and the single collector efficiency, η_T, which follows from straightforward geometric and mass balance consider ations:

$$\lambda_0 = \frac{3(1-\varepsilon)}{2d_C}\eta_T \qquad (7.15)$$

where ε is the porosity of the filter bed (=voids volume/total bed volume).

Using Equations (7.14) and (7.15), we can calculate the percent removal of particles in a filter bed of defined characteristics. If the bed depth is 1 m and all other properties are the same as for Figure 7.12, the results in Figure 7.13 are obtained. The effect of increasing flow velocity is clearly shown. At 1 m/h-particle removal is close to 100% throughout the entire size range, whereas at 20 m/h, removal of particles in the 0.5–1 μm range is little more than 50%. These calculations have assumed that particles attach to collectors at first contact (α=1). For smaller collision efficiencies, the collector efficiency η_T, and hence the filter coefficient λ_0, would be correspondingly lower, giving less removal in the filter.

Equation (7.14) indicates an exponential decline in filtrate concentration with depth, which means that the upper layers of the filter bed remove more particles than lower layers. These particles coat the filter grains, so the filter coefficient departs from the "clean bed" value λ_0. Initially, the filter coefficient may increase as a result of deposited particles, giving improving performance in the early stages (a process known as *filter*

ripening). However, increasing deposits eventually cause the filter coefficient to decrease, partly

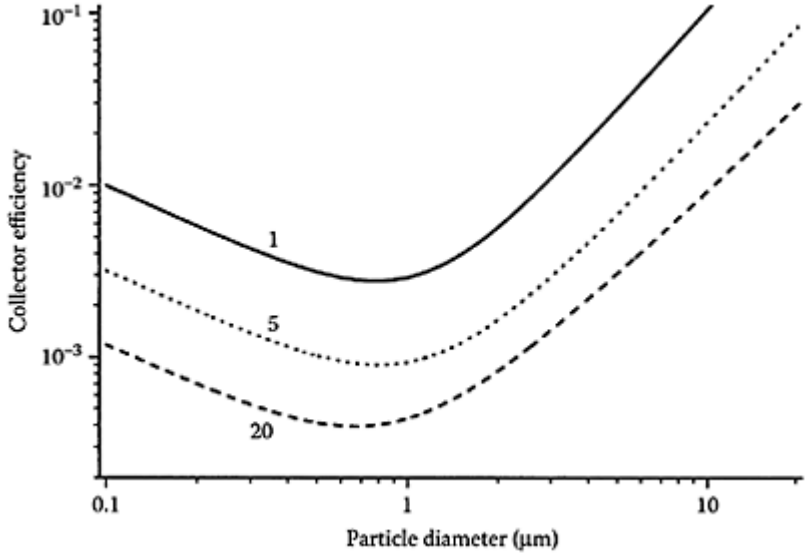

Figure 7.12 Total collector efficiency at different flow rates (values shown on curves as approach velocities in m/h). Other conditions are as for Figure 7.11.

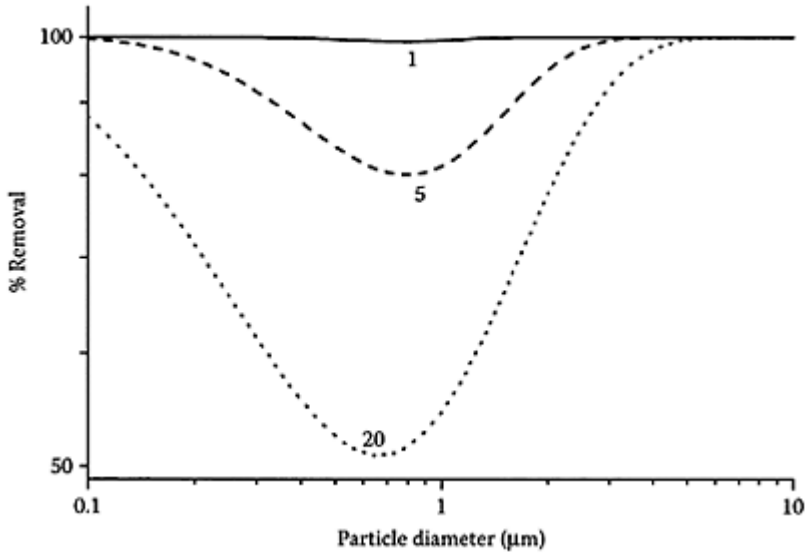

Figure 7.13 Showing the effect of flow rate (approach velocity in m/h) on % removal of particles by a filter of 1-m depth, as a function of particle size. Other conditions are given in the text.

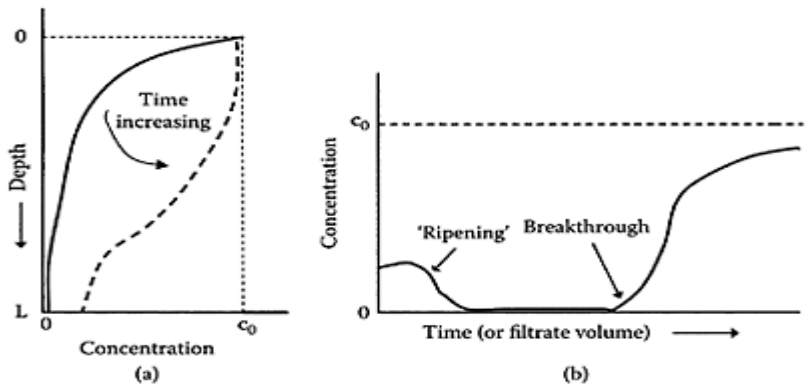

Figure 7.14 (a) Particle concentration versus depth in a granular filter; (b) Particle concentration in the filtrate as a function of time (or filtrate volume), showing an initial "ripening" phase and then breakthrough (not to scale).

because, as deposits build up, the pores become increasingly clogged and so the head loss rises and the local flow velocity increases (assuming that the volume flow rate is constant). This makes prediction of the performance of real filters difficult.

Figure 7.14(a) shows the particle concentration in water flowing through a deep bed filter at two stages. Early in the filter run, the bed is still clean and the profile is exponential, as predicted by Equation (7.14). Later, very little removal occurs in the upper layers, as shown by the unchanging particle concentration in this region. A plot of the amount of deposited particles in the bed would show similar profiles. Eventually the particle concentration in the filtrate begins to rise, as the deposits penetrate further into the bed and *breakthrough* occurs, as shown in Figure 7.14(b). This shows an initial decrease in filtrate concentration (the *ripening* effect mentioned earlier), then a period when the concentration remains at a nearly constant low value, and finally a significant rise, ultimately to the inlet concentration, when no removal occurs.

When the filtrate concentration reaches some predefined limit, the filter run has to be terminated. Another reason is that the overall head loss may rise to a value such that the desired flow rate can no longer be maintained. In both cases the filter needs to be cleaned to remove accumulated deposits; this is done by *backwashing*, which involves flowing clean water upward through the bed, usually causing *fluidization* of the grains. In practice, back-washing is often carried out on a periodic basis—say, every 2 days—rather than as a result of breakthrough or head loss buildup.

As mentioned earlier, deep bed filters do not provide absolute filtration. Where complete removal of particles more than a certain size is needed, then some form of *membrane filtration* can be used.

Table 7.1 Typical Characteristics of Membrane Processes

Process	Operating Pressure (Bar)	Pore Size (nm)	Molecular Weight Cut-Off Range	Size Cut-Off Range (nm)
Microfiltration	<4	100–3000	>500,000	50–3000
Ultrafiltration	2–10	10–200	1,000–1,000,000	15–200
Nanofiltration	5–40	1–10	100–20,000	1–100
Reverse osmosis (hyperfiltration)	15–150	<2	<200	<1

7.5.2 Membrane filtration

Membrane filtration is a separation process that uses thin polymer or ceramic membranes with pores in a certain size range. The finer the pores, the smaller the particles removed. Polymer materials used for membrane manufacture are, for instance, cellulose acetate and polysulfone. Actual polymer filter membranes are very thin (typically 0.1–1 μm) and supported by a porous structural substrate. Ceramic membranes are usually manufactured from metal oxides, such as alumina, often using some form of sol-gel process. We shall

not go into detail of membrane manufacture, but we will outline some of the important features of membrane filtration.

Membrane filtration can be divided into four categories, depending on the effective pore size of the membrane and hence the size of the impurities removed. In order of decreasing pore size, they are as follows:

1. Microfiltration (MF)
2. Ultrafiltration (UF)
3. Nanofiltration
4. Hyperfiltration (or reverse osmosis)

Table 7.1 summarizes the essential features of these processes, such as pore size and operating pressure, although these are not absolute values and different ranges may be quoted elsewhere. The smaller the pore size, the larger the pressure needed to drive water through the membrane at an acceptable rate. Also shown are the particle size cut-off ranges and similar information in terms of molecular weight because membranes can also remove dissolved impurities, depending on the effective size of their molecules. In some listings, particle size cut-off is only quoted for microfiltration and molecular weight ranges are given for the other processes. (The term *molecular weight cut-off*, MWCO, is often used.)

Reverse osmosis (hyperfiltration) is different from other types of membrane filtration. The latter require pressure to overcome the hydraulic resistance of the membrane, whereas reverse osmosis, as the name implies, has to overcome the natural osmotic pressure of the solution. Reverse osmosis membranes have pores so small that they reject most solutes, including dissolved salts, but allow water to pass, provided that sufficient pressure is applied. To some extent, similar remarks apply to nanofiltration, which is an

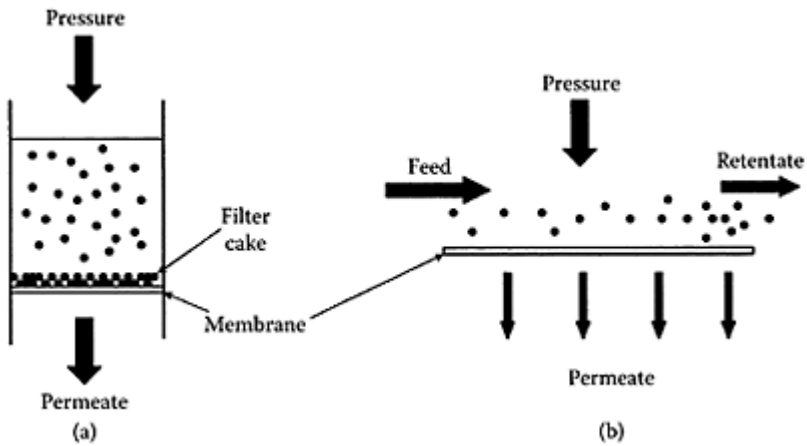

Figure 7.15 Two forms of membrane filtration: (a) Dead-end and (b) Cross-flow.

intermediate case between hyperfiltration and UF. Reverse osmosis is an attractive process for the desalination of brackish waters and sea water. In the latter case, the osmotic pressure is about 28 bar, so that considerably higher pressures (more than 50 bar) have to be applied in practice. Because reverse osmosis is mainly used to remove solutes from water and we are concerned with *particle* separation, most attention will be paid to the other processes.

There are two ways of operating a membrane filter:

- *Dead-end* filtration
- *Cross-flow* filtration

Both of these are widely used, and they are illustrated schematically in Figure 7.15.

In dead-end filtration, all of the feedwater flows through the membrane (as *permeate*), so that all impurities that are too large to pass through the pores accumulate in the filter module. Some means of removing these is necessary.

Cross-flow filtration involves flowing the feedwater parallel to the membrane surface, with only a proportion passing through the membrane. The retained impurities remain in the *retentate*, which is normally recirculated. This mode of operation inevitably generates a liquid waste stream containing the removed impurities at high concentration. However, a membrane filter operated in dead-end mode needs to be cleaned periodically (see later), and this also produces a concentrated waste.

Practical membrane filters can be packaged in various ways:

- Tubular membranes, with an internal diameter greater than 3 mm, arranged in a bundle
- Capillary tube bundles, with hundreds or thousands of individual hollow fibers, with internal diameters of the order of 1 mm, bundled into a compact module
- Spiral wound modules, consisting of membrane sheets wound around a central porous tubular core, which acts as a permeate collector
- Plate and frame modules, in which parallel flat membrane sheets are stacked, with porous spacers, which serve to collect the permeate

These arrangements are all intended to pack a large filtration area into a small space because the flux through a membrane may be small. The flux, often expressed in L m^{-2} h^{-1}, depends on the membrane type (especially pore size) and the *transmembrane pressure* (TMP). Typical values are around 150 L m^{-2} h^{-1} bar^{-1} for microfiltration (MF) membranes. Ultrafiltration and nano-filtration membranes have much lower values and need correspondingly higher pressures to achieve comparable flux values.

Assuming that an MF unit operates at a TMP of 1 bar and achieves a flux of 150 L m^{-2} h^{-1}, then the corresponding approach velocity is only 0.15 m/h, which is much lower than the value for rapid sand filtration (5–30 m/ h). Consequently, a much higher filtration area is needed.

The most significant problem with all membrane filters is *fouling* of the membranes. This is the process whereby contaminants from the feedwater deposit on the membrane surface or within the pores and cause a decrease in flux (or an increase in TMP to maintain the flux). Some of the retained material may be easily removed by washing, whereas some may become fixed to the membrane in a way that makes removal difficult, if not impossible. These give *reversible* and *irreversible* fouling, respectively.

Fouling always occurs to some extent in dead-end filtration because the impurities have to deposit on the membrane to be removed. As particles approach the membrane surface, they may enter pores and cause blockage or form a layer of deposit, as shown in Figure 7.15. The deposited layer is sometimes known as a *filter cake* because it can itself act as a filter for approaching particles. The filter cake can also add considerably to the overall resistance to flow, often giving several times more resistance than the underlying membrane. The formation of a filter cake causes a steady increase of TMP to give a certain flux. On removing the cake by washing the filter, the TMP may return to the original clean membrane value or, more likely, show only partial recovery, which indicates some irreversible fouling. This is shown schematically in Figure 7.16.

Cross-flow filtration is inherently less likely to give fouling problems because impurities are carried parallel to the membrane surface. However, some deposition on the membrane can occur, giving a gradual flux decline or increase in TMP. It has been shown that keeping the transmembrane flux below a certain value (the *critical flux*) greatly reduces the fouling problem. The critical flux depends on such parameters as particle size and shear rate at the membrane surface.

Membrane-fouling problems can be significantly reduced by some form of pretreatment of the water. For instance, coagulation/flocculation can be effective because the increased size of particles (flocs) gives less penetration into pores and more permeable filter cakes.

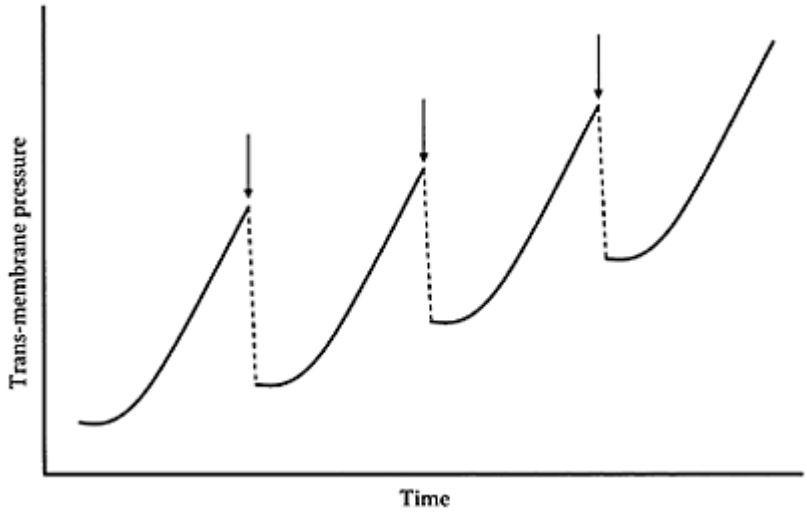

Figure 7.16 Variation of transmembrane pressure (TMP) with time during dead-end filtration (for constant flux). Pressure increases as fouling occurs (buildup of filter cake). Arrows indicate membrane cleaning.

Note that there is only a partial recovery of permeability after each cycle, indicating some *irreversible fouling*.

When a membrane has become fouled, it has to be cleaned in some way. Available methods include flushing with water in either a forward or reverse direction. The latter *(back flushing)* is generally more effective, although not for impurities strongly adsorbed on the membrane. Cleaning may be considerably improved by the application of vibration (up to 1000 Hz) or ultrasound (around 40 kHz). Finally, a range of chemical cleaning methods can be used.

For particle removal in water and wastewater treatment, the choice is usually between MF and UF. These operate under similar conditions, apart from the different TMP required. Because of the smaller pores UF membranes are better able to remove small, colloidal particles, including all viruses. MF membranes only give partial virus removal. UF membranes can also give significant removal of dissolved organic material, such as humic substances.

Further reading

Boller, M. (Ed.), *Nano and Microparticles in Water and Wastewater Treatment*, IWA Conference, Zurich, September 22–24, 1993. (Published in *Water Sci. Technol*, 50(12), 2004.)

Casey, T.J., *Unit Treatment Processes in Water and Wastewater Engineering*, Wiley, Chichester, 1997.

Haarhoff, J. and Edzwald, J.K., Dissolved air flotation modelling: insights and shortcomings, *J. Wat Suppl: Res. & Technol—AQUA*, 53, 127, 2004.

McEwen, J.B. (Ed.), *Treatment Process Selection for Particle Removal*, American Water Works Association, Denver, 1998.

Svarovsky, L. (Ed.), *Solid-Liquid Separation*, Butterworth-Heinemann, Oxford, 2000.

Index